PROPERTY OF
LIBRARY

Vol. 33, No. 2 August, 1943

Journal of Experimental Psychology

EDITED BY

SAMUEL W. FERNBERGER, ON LEAVE
FRANCIS W. IRWIN, ACTING EDITOR
UNIVERSITY OF PENNSYLVANIA

CONSULTING EDITORS

CHARLES W. BRAY, *Princeton University;* ELMER K. CULLER, *University of Rochester;* CLARENCE H. GRAHAM, *Brown University;* JOY P. GUILFORD, *University of Southern California;* FRANCIS W. IRWIN, *University of Pennsylvania;* DONALD G. MARQUIS, *Yale University;* ARTHUR W. MELTON, *University of Missouri;* CARROLL C. PRATT, *Rutgers University.*

CONTENTS

Factors Influencing the Time Error in Judgments of Visual Extent: D. C. McCLELLAND 81

Bone Conduction During Experimental Fixation of the Stapes: K. R. SMITH .. 96

The Relative Difficulty of Morse Code Alphabet Characters Learned by the Whole Method: S. D. S. SPRAGG 108

An Experimental Comparison of Code-Learning Methods: J. P. SEWARD ... 115

The Occurrence of Plateaus in Telegraphy: H. B. REED AND H. A. ZINSZER .. 130

Generalization in the Initial Stages of Learning Nonsense Syllables: II. Partial and Inadequate Responses: B. R. PHILIP AND H. E. PEIXOTTO 136

The Influence of Size of Test Stimuli, Interpupillary Distance, and Age on Stereoscopic Depth Perception: L. C. MEAD 148

A Cumulative Graphic Work-Recorder: O. H. MOWRER 159

PUBLISHED MONTHLY, TWO VOLUMES PER YEAR
BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879

PUBLICATIONS OF

THE AMERICAN PSYCHOLOGICAL ASSOCIATION

WILLARD L. VALENTINE, *Business Manager*

PSYCHOLOGICAL REVIEW

HERBERT S. LANGFELD, *Editor*
Princeton University

Contains original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.
Subscription: $5.50 (Foreign, $5.75). Single copies, $1.00.

PSYCHOLOGICAL BULLETIN

JOHN E. ANDERSON, *Editor*
University of Minnesota

Contains critical reviews of books and articles, psychological news and notes, university notices, and announcements. Appears monthly (10 issues), the annual volume comprising about 665 pages. Special issues of the BULLETIN consist of general reviews of recent work in some department of psychology.
Subscription: $7.00 (Foreign, $7.25). Single copies, 75c.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

S. W. FERNBERGER, *Editor, on Leave,* FRANCIS W. IRWIN, *Acting Editor*
University of Pennsylvania

Contains original contributions of an experimental character. Appears monthly (since January, 1937), two volumes per year, each volume of six numbers containing about 520 pages.
Subscription: $14.00 ($7.00 per volume; Foreign, $7.25). Single copies, $1.25.

PSYCHOLOGICAL ABSTRACTS

WALTER S. HUNTER, *Editor*
Brown University

Appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.
Subscription: $7.00 (Foreign, $7.25). Single copies, 75c.

PSYCHOLOGICAL MONOGRAPHS

JOHN F. DASHIELL, *Editor*
University of North Carolina

Consist of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The MONOGRAPHS appear at irregular intervals and are gathered into volumes of about 500 pages.
Subscription: $6.00 per volume (Foreign, $6.30).

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

GORDON W. ALLPORT, *Editor*
Harvard University

Appears quarterly, January, April, July, October, the four numbers comprising a volume of 560 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.
Subscription: $5.00 (Foreign, $5.25). Single copies, $1.50.

JOURNAL OF APPLIED PSYCHOLOGY

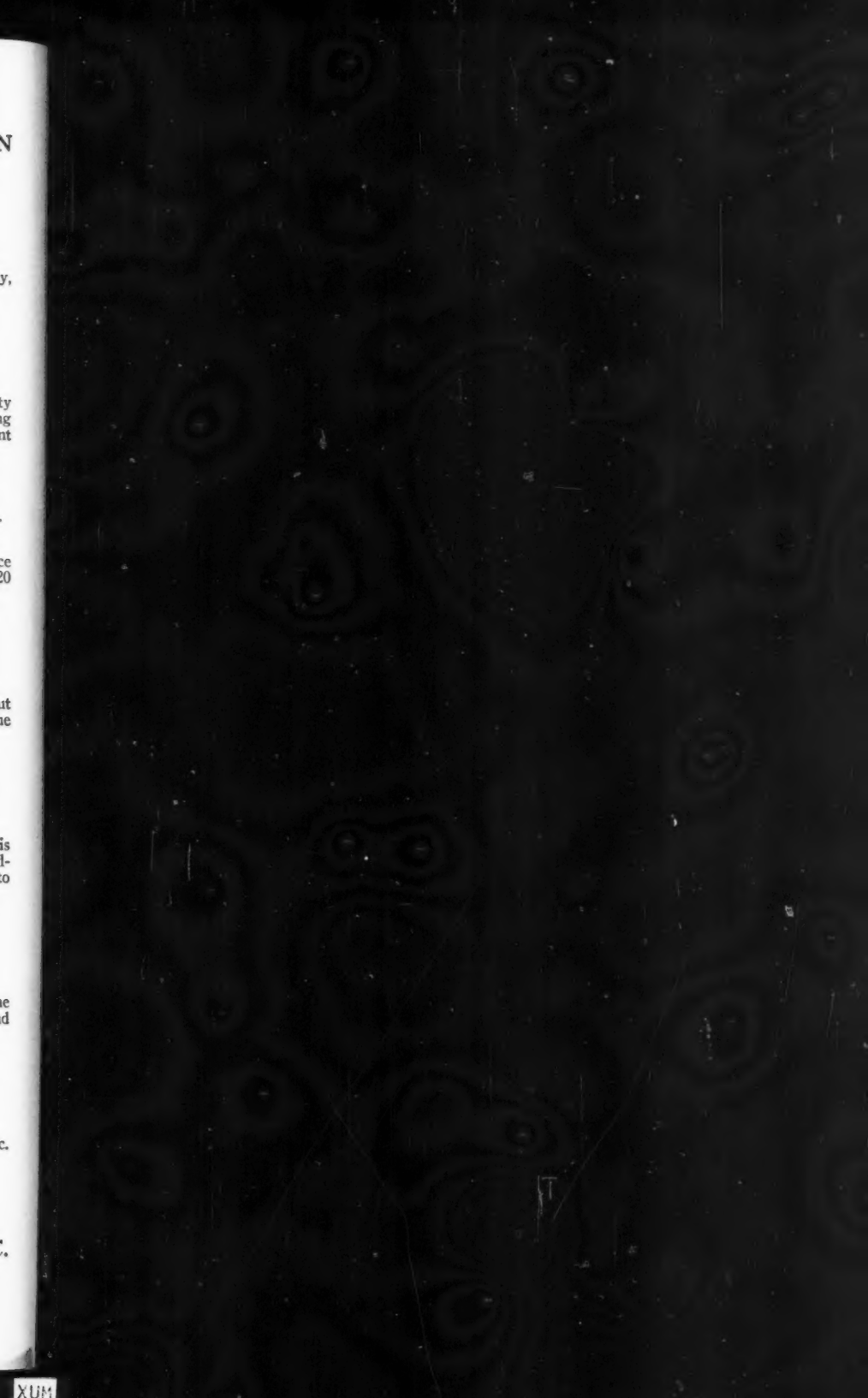
DONALD G. PATERSON, *Editor*
University of Minnesota

Covers the applications of psychology in business, industry, education, government, etc. Appears bi-monthly, February, April, June, August, October, and December.
Subscription: $6.00. Single copies, $1.25.

Subscriptions, orders, and business communications should be sent to

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS



XUM

Journal of
Experimental Psychology

VOL. 33, NO. 2 AUGUST, 1943

FACTORS INFLUENCING THE TIME ERROR IN JUDGMENTS OF VISUAL EXTENT

BY DAVID C. McCLELLAND

Wesleyan University

Marchetti (**12**) has reported a positive time error in judgments of visual extent, a result which, as he points out, is in direct contradiction to the negative time error found in nearly all other comparisons by the method of constant stimuli, when the stimuli are separated by intervals longer than three seconds. Conventional theories (**9**, **18**, **22**) of the negative time error have attributed it either to assimilation or to the fading trace of the first stimulus. At first glance Marchetti's result appears to contradict both theories. At any rate it demands some explanation.

An explanation in terms of fading trace does not readily present itself, but one in terms of assimilation does. The stimuli used by Marchetti were always black lines drawn on a white piece of cardboard. Consequently, the lines always appeared against longer background contours. By the method he used there was no possible way in which the contours could be shorter than the lines; so the standard line may have been assimilated to the longer contours of the background on which it was presented, thus becoming itself longer and causing a preponderance of shorter judgments of the comparison line. The positive time error which should and did result would have the same sort of explanation as the negative time error which results in other comparisons, such as that between auditory intensities when the standard stimulus is usually presented on a background of lesser magnitude (**22**).

It was with this hypothesis in mind that the present series of experiments was designed. If the hypothesis is correct, the positive time error should disappear if the lines are without a background, or at least without a background containing linear contours to which the standard could assimilate. The lines were therefore projected

XUM

on a screen in a darkened room, appearing as lines of light on a dark contourless surface. It was further predicted that if this procedure succeeded in destroying the positive time error, the latter would reappear if the room were illuminated enough to make the contours of the screen on which the lines were projected visible.

Procedure

The lines were made by cutting slits in black paper and mounting the paper in standard glass slides. They could then be thrown on the screen by means of an ordinary slide-projector, and appeared as lines of light approximating film rather than object color. The room used for the experiment was a large lecture room, 44 feet long by 34 ft. wide, and 19 ft. deep at the end which held the screen on which the lines were projected. The slide projector was in the back of the room about 38 ft. from the silver screen which was itself a very large rectangle, 18 ft. long by 9 ft. high, permanently fixed to the wall for use as a movie projection screen. The subjects were seated in rows at distances from the screen varying from around 15 to 30 ft. The lines were projected at a height of 10 ft. above the floor, *i.e.*, about 5 ft. above sitting eye-level for the nearest subjects and about at eye level for the subjects seated farthest away, since each row of seats was raised a foot above the row just in front. A record was kept of the sitting location of each subject in case this should in some way affect the results.

The method of constant stimuli was used with a standard line 13 in. long and 5 comparison stimuli 11, 12, 13, 14 and 15 in. in length. In some experiments the standard was 11 in. long in which case the comparison stimuli were 9, 10, 11, 12 and 13 in. long. Standards about a foot long were chosen because it was thought possible that the subjects might have an absolute standard of this length.

After a ready signal the standard stimulus was presented for 2 sec. and after a 7–8 sec. pause the comparison stimulus appeared, also for 2 sec. The subjects were to write down immediately whether the second line appeared longer than, equal to, or shorter than the first line. They were run in groups of from 11 to 39 at a time; so they were instructed, after being informed of the procedure, not to be influenced by the judgments of others. They were also told that all that was desired was their immediate impression of the length of the second line relative to the first, and that they were not to try and reason out whether the second should be longer or shorter than the first. They were completely unaware of the purpose of the experiments until the experiments were completed, and probably thought accuracy of discrimination was being tested. They were all undergraduate women between the ages of 17 and 22, registered in psychology courses at Connecticut College.

The comparison stimuli were presented in a predetermined random order such that each appeared once in each successive group of 5 comparisons and such that the extent of increases and decreases in length of successive comparison stimuli balanced out in case there should be any serial effects. The standard was always presented first but many subjects were not aware of this as later questioning proved. Usually the subjects made 15 judgments, 3 for each comparison. The judgments were classified into the shorter, equal and longer categories and treated by the Spearman summation method (22, p. 402).

Variations in the procedure will be noted under each condition in discussing results.

Results

All the results obtained under different conditions are brought together in Table I for purposes of easy comparison. The columns on the left describe the conditions of the experiment, the one in the middle the experimental variable, and the ones on the right the results in terms of points of subjective equality.

The first column in Table I gives the letter by which each experiment will be referred to. The experiments were performed in the

TABLE I

SUMMARY OF EXPERIMENTAL CONDITIONS AND RESULTS

1	2	3	4	5	6	7	8	9	10
C	*Cl*	*N*	*n*	*O*	*St.*	Field Condition	*PSE'*	*PSE*	σ_m
A.	*X*	35	105	1	13″	Dark	12.81	12.83	.07
B.	*X*	39	117	2	13″	Dim	12.72	12.83	.06
C.	*Y*	21	63	2	11″	Dim	10.67	10.70	.09
D.	*Z*	19	95	2′	11″	Dim, *St.* framed	11.03	11.00	.06
E.	*Y*	24	110	2′	13″	Dim, both framed	12.84	12.93	.06
F.	*X*	37	111	2	13″	Dim, both framed, fixation frame	12.85	12.92	.06
G.	*Y*	16	80	2′	13″	Dim, constant frame	12.90	12.96	.07
H.	*X*	32	32	–	13″	Dim, oblong constant frame	(I.P. = −44)		
						St. in oblong frame	(I.P. = +16)		
J.	Z	11	33	1	13″	Dark, light source interrupted	13.10	13.14	.11
K.	Z	11	33	1	13″	Dim, light source interrupted	12.97	13.00	.13

1. Condition designation.
2. Class of subjects.
3. Number of subjects in class.
4. Number of judgments for each pair compared (n).
5. Order of presentation.
6. Length of the standard line in inches.
7. Field condition, experimental variable.
8. *PSE'*: Point of subjective equality as the average of the upper and lower limens determined by summation method.
9. *PSE* as the average of the equal judgments.
10. Standard errors of the *PSE*s in column 9.

I.P. = Index of preponderance as determined by the formula:

$$\frac{\text{Shorter} - \text{Longer}}{n} \times 100.$$

order listed, all within a period of about 6 weeks. The second column uses the letters *X*, *Y* and *Z* to refer to three different classes in psychology containing entirely different subjects. The third column, labelled *N*, records the number of subjects used in the experiment and the 4th column, labelled *n*, the total number of judgments for each comparison. The number of judgments each subject made is not listed but can be found by the formula $\frac{5n}{N}$. The fifth column, labelled *O*, refers to the order of presentation of the comparison stimuli. Order 2 is the converse of order 1 (that is, if the 15 in. comparison stimulus was presented first in order 1, the 11 in. one would appear first in order 2, etc.). Order 2′ is just like order 2 through the first 15 comparisons. After that 10 more comparisons were made in a random order (or 15 in the case of Experiment G).

Column 6, labelled *St.*, refers to the length of the standard lines, and column 7 lists the important conditions of the field which differed from experiment to experiment.

Columns 8, 9 and 10 give the major results of the experiments. In column 8 are given the points of subjective equality as determined by averaging the upper and lower limens obtained by the Spearman arithmetic mean method of computation (**22**, p. 402). There appears to be no way in which the variability of this *PSE* can be estimated (*cf.* **4**) without taking into account the correlation between the two limens.[1] Since this correlation is not known and cannot be calculated except by fractioning the data, and at the expense of considerable labor, another method of determining the *PSE* was adopted which yields a good measure of its variability. In column 9 are presented the *PSE*s when defined as the means of the distributions of equal judgments. Culler (**3**, p. 71) specifically warns against using any measure of central tendency from the equal category on the grounds that it bears no relation to the relative magnitudes of the 'greater' or 'less' categories. He was interested however in establishing the exact point where the judgment of 'greater' was equally probable to the judgment of 'less.' On logical grounds it would seem just as proper to define the *PSE* as that value which, on the average, was judged equal to the standard. Furthermore, it should be remembered that the purpose of these experiments is not to determine the *PSE* as accurately as possible, but rather to obtain a value which will permit valid comparisons between *PSE*s under different conditions. The requirement of such a value is not absolute accuracy but equality of accuracy under different conditions. Finally it should be noted that the values in column 8 and column 9 do not differ very much, for the most part. Because of all this, and particularly in view of the fact that the variability of the mean of the equal judgments can be readily estimated from the usual formulae, the *PSE* as determined in column 9 will be used throughout the subsequent discussion.

It will be noted that the time error can be obtained in each experiment by subtracting the length of the standard from the length of the *PSE* (*PSE–St.*). (A *PSE* less than the standard represents by convention a negative time error.) Previous studies in this field have used variability statistics only to determine whether the time error is reliably positive or negative (**6**). Such a procedure would not be of any use here, because the problem is not so much the direction of the time error as whether it differs significantly under one set of conditions from what it is under another. It seems perfectly logical to test the reliability of differences in *PSE*s, although the usual technique has rarely if ever been applied to psychophysical data. Perhaps this is because in psychophysics the interest has been in determining as accurately as possible individual limens (checked for idiosyncrasy by using two or three subjects), rather than in group limens and group differences. In the present experiment, group limens and group *PSE*s have been determined, not because they have any particular psychophysical meaning, but because they are representative numbers which will be influenced in various ways by changes in conditions. In other words the attempt has been to make psychophysical methods yield group scores which may be compared just as group averages obtained in other fields conventionally are.

[1] According to F. W. Irwin in a communication to the writer the formula would be:

$$PE_{PSE} = .707\, PE_L \sqrt{1 + r_{L_1 L_2}} \quad \text{if} \quad PE_{L_1} = PE_{L_2}.$$

There is a question as to whether the standard errors of the means of the equal judgments as represented in column 10 are proper estimates of the variability of those means. The cut-tail error in each case is small enough to be neglected, since the equal judgments falling in the two extreme categories constituted in every instance less than 5 percent of the total judgments. This should lead to a very slight underestimation of the variability of the *PSE*s. Since the logic of the matter is not altogether clear, it was decided to test empirically, by fractionating the data, whether an obtained distribution of sub-means differed from the average of the estimates of the variability of those means. The ratio of these two values, $\sigma_{obt.}/av.\ \sigma_{est.}$, the Lexian ratio described by Culler (3), Melton (13) and others, was 1.278 for condition A, when the 35 subjects were divided into seven sub-groups of five each. The deviation from 1.00 is just 1.22 times its *PE* (obtained by the formula $PE_L = .4769L/N$), indicating that it would frequently arise by chance. Whatever the logic of the case, the standard error of the mean of the equal judgments does estimate rather well how that mean will vary.

Finally, it should be noted that the estimates of variability are valid only for *PSE*s determined under similar conditions. The conditions are exceptional in that predictions are made neither for individuals nor for groups of individuals, since each individual contributes more than one score to the average. This fact does not detract, however, from the comparisons between *PSE*s, since they were all computed in the same way.

A. In the first experiment, which was designed to discover the direction of the time error in a contourless field, heavy black curtains were drawn over the windows. The lights were turned out before the presentation of the standard stimulus and remained out until after the presentation of the comparison stimulus; they were then turned on again to permit the subject to record his judgment and also to prevent dark adaptation. The room was not totally dark. Some light came in from the hall through the rectangular transom in the door, which was to one side and considerably below the place where the lines appeared. But aside from this, the room was free of contours to which the standard could assimilate in the darkness. The result ($PSE = 12.83$) is a reliably negative time error in direct contradiction to Marchetti's results (12). The chances that the *PSE* would fluctuate up to a value of 13.00 are less than 2 in 100, (C.R. = 2.43).

B. The hypothesis was that if the illumination were now increased, the standard would assimilate to the contours in the background and produce a positive time error like Marchetti's. Instead, under these conditions the negative time error persisted. The *PSE* is the same (12.83) as under the previous condition. It is possible that the order of presentation influenced the result, but subsequent comparisons in which it was varied again make this doubtful.

C. The results under B contradicted the original hypothesis and necessitated a new one. It was thought possible that the standard was assimilated down to an absolute standard of around 12 in., making for an increase in the judgments of longer or a negative time error. So a standard shorter than 12 in. was chosen, on the supposition that assimilation now to an absolute standard a foot long would cause an increase in the standard in the interval between comparisons, and, consequently a positive time error. Nothing of the sort

occurred. If anything, the negative time error increased again. The chances that the *PSE* of 10.70 would fluctuate up to a value equal to the standard are less than 1 in 100 ($P < .01$).

D. The hypothesis was next advanced that the large negative time errors so far obtained were due to the fact that the contours of the screen were so large and so remote from the lines projected that they were not sufficiently similar or near for assimilation to take place. Consequently, the standard was enclosed in a 33-in. square frame of light lines of the same width as the standard and made in the same way by cutting slits in the black paper enclosed in the slide. The hypothesis was that if assimilation were a factor at all, the standard should assimilate to its larger frame and cause an increasing proportion of shorter judgments. The hypothesis was confirmed. The large negative time error was wiped out and the *PSE* (11.00) equalled the standard. The critical ratio of the difference between the *PSE* in Experiments C and D is 2.78, $P < .01$.

E. But the conditions of D are unlike those of Marchetti's (12). Both his standard and comparison stimuli were framed. When this condition is reproduced here, the negative time error still appears (*PSE* = 12.93), but it is not as large as when both are unframed. The 13-in. standard was used again because from a comparison of the time errors in B and C the size of the standard appeared to have had no important effect.

F. The negative time error was still unexplained. Perhaps it is due to the fact that in Marchetti's experiment there was a fixation field between the presentation of the standard and comparison and here there was not. To duplicate this condition a solid white 33-in. square was interpolated between the two stimuli. This had no noticeable effect. The *PSE* was about the same as in E (12.92).

G. An attempt was made to approximate the tachistoscopic conditions which gave rise to a positive time error for Marchetti. The lines were presented in the 33 in. frame, but the frame remained on the field continuously, instead of appearing and disappearing as in Experiments D, E and F. As in Experiments E and F, the negative time error (*PSE* = 12.96) was reduced as compared with what was obtained without any frames, but not reliably so. Since in Experiments E, F and G both standard and comparison were framed, and since the differences that existed among them hardly changed the *PSE*s at all, it seems permissible to combine the three distributions of equal judgments and to compare the mean of the total with the mean of Experiments A and B similarly combined when no frames were used. If this is done the difference between the *PSE*s is .105 (12.934 − 12.829) and its standard error .0575. The critical ratio is 1.83, indicating that the chances are less than 5 in 100 that a dif-

ference that large and in that direction could have arisen by random sampling. This means that introducing the frames decreased the negativity of the time error, as would be expected according to an assimilation hypothesis. The fact remains, however, that the time errors are not positive as in Marchetti's experiments. So the search for the difference in the conditions continued.

H. Although some of Marchetti's backgrounds were square, under certain conditions the lines appeared on rectangular white cards (12). Perhaps this was the factor that produced his positive time error. Instead of running through the whole series, two objectively equal stimuli were presented, both in the center of a rectangle just as long as the 33-in. square but only half as high. The index of preponderance [2] is − 44 indicating that the time error is still negative. When the standard appeared in the rectangle and the comparison did not, the I.P. became positive (+ 16) as in Experiment D.

At this point it was thought desirable to obtain some estimates of the absolute lengths of the lines. On this day, before any lines were shown, the subjects were asked to write down an estimate in inches of the length and width of the first line in each of the comparisons they had made just one week previously. Some subjects did not know that a line of the same length was always presented first. They wrote down the best estimate they could make anyway. The fact that the standard was doubtless mixed up in memory with the comparison stimuli could not influence the estimate in any consistent way, since shorter and longer comparison stimuli appeared equally often. After this memory estimate, they made the judgment of the length of two equal lines which appeared in a rectangle. Then they were told that the first line they had seen was the same length as the one whose length they had just estimated from memory. Now, some 30 seconds after they had seen the line, they were asked to estimate its length and width again. Finally they were presented with the same line and were allowed to take as long as they liked to estimate its length and width while they were looking at it.

The results in Table II show rather clearly that the longer the time since the subjects had seen the line the larger they estimated it to be. The memory trace of the line does appear to get larger with time, a fact which is consistent with Marchetti's results (12) and which would yield the positive time error he obtained. The difference between the first two estimates of absolute length, 1.70 in., is 2.24 times its standard error (calculated by the direct method), indicating that the difference would rarely ($P < .03$) have arisen by

[2] The index of preponderance was calculated by the formula given by Marchetti (12):

$$\frac{\text{Shorter} - \text{Longer}}{\text{Number of judgments}} \times 100.$$

chance. A similar comparison between the second and third estimates is not altogether justified, because the line had appeared inside a rectangle in the second estimate and without any frame for the third estimate. The result may, therefore, simply reflect assimilation of the second estimate as compared with none for the third.

TABLE II

MEAN ESTIMATED SIZE (IN INCHES) OF A 13-INCH LINE SEEN AT DIFFERENT TIMES PREVIOUSLY

Time Between Presentation and Judgment	Estimated Length (in inches)	Estimated Width (in inches)
7 days	13.76	.96
30 seconds	12.06	.72
None	10.85	.63

The important fact here is that the line judged from immediate memory is reliably shorter than the line judged from remote memory. It is true that the line in remote memory was presented in a square, rather than in a rectangle, but if this influenced the results in any way it would tend to make the difference smaller, since the remote assimilation (to a square) would be less than the immediate (to a rectangle). Although the conditions for these estimates are not ideal, the results do indicate that positive time errors are to be expected at least so far as the memory trace of the line is concerned.

J. So far the evidence is somewhat paradoxical. The absolute judgments indicate along with Marchetti's experiments that positive time errors should be obtained, rather than the consistently negative ones actually obtained. At this point it was observed that the method of presenting the stimuli might be the cause of the negative time errors. The light in the projector which threw the line on the screen was simply turned on for two seconds and then turned off. Since the bulb was the ordinary filament type, this meant that it cooled off gradually. Consequently the line faded from view and did not disappear abruptly. This appeared to be a major departure from Marchetti's procedure, and might easily have provided the fading standard which produced the negative time errors. It had been supposed up to this time that, although the line faded, this would not necessarily mean that it got shorter; but the supposition was tested and found mistaken.

Condition J is just like Condition A, except that the light source was cut off sharply by an intercepting opaque object. Further, the subjects were asked to close their eyes between presentations so that there would be no possibility of assimilation from the visible contours in the room. Under these conditions the time error changed

from $-.17$ to $+.14$ (C.R. = 2.38, $P < .02$), providing rather conclusive evidence that it was the fading stimulus source which had been causing negative time errors up to this point.

K. Condition B was now repeated with the light source interrupted to test the hypothesis that assimilation to the background would increase the positive time error. Instead the time error (.00) shifts in the opposite direction though not reliably so.[3] The *PSE*s in column 8 show the same tendency in A and B.

Discussion

Fading Stimulus Source.—The last two experiments throw a new light on the whole series and make the results obtained intelligible, particularly when compared with Marchetti's (**12**). The fading stimulus source was responsible for the negativity of the time errors throughout. The standard was subject to an apparent contraction in length due to fading, a change discovered by Korte in 1916 and named by him 'gamma-movement' (**1**).

None of the time errors obtained are reliably positive as Marchetti's appear to be, but it is very likely that they would have been under Conditions E, F and G (framed stimuli), the conditions most like his, had the stimulus source been interrupted instead of turned off. These experiments would then tend to support, and certainly not to contradict, Marchetti's result that the time error is 'normally' positive in judgments of visual extent, a finding which is also consistent with the absolute estimates from remote and immediate memory.

The failure to control adequately the presentation of the initial stimulus, which at first appeared to be an error in technique, on second glance suggested that other experiments of this type which have uniformly reported negative time errors might be open to the same criticism. In fact it should be extremely difficult, outside of the field of vision, to control the sensory input resulting from the standard stimulus so that it will not fade off very gradually leaving a weak final impression with which the second stimulus is compared as soon as it is presented and before it has a chance to fade.

Fading in the sensory input can be due logically either to lag in the cessation of the physical stimulus or to gradual receptor adjustments following stimulation. If it can be shown that the negative time errors obtained can be attributed to either of these causes, it will no longer be necessary to postulate a fading memory trace, as Köhler does (**8**), or a change in set (**22**) or a change in any central factor. In fact, the whole phenomenon would then turn out to be

[3] It should be noted that Experiments J and K were done on the same day in one continuous session with the same group of subjects. Five trials of J were alternated with 5 trials of K in the order ABBABA.

an artifact of the method of presentation and the characteristics of the receptor mechanism, and would have nothing to do with memory at all.

An investigation of the major fields for which negative time errors have been reported indicates that such might easily be the case. In the early experiments on auditory intensities made by Köhler (**8**) and Lauenstein (**9**), which gave rise to much of the present interest in the phenomenon, no attempt was made apparently to control the cessation of the physical stimuli used, which were for the most part short clicks made by a relay but presented by means of a loudspeaker in non-sound-screened rooms, sometimes even in an ordinary lecture room (**9**, p. 158). Under such conditions the sound will continue reverberating a few tenths of a second after the current stops; and what is more important, it will diminish in intensity only gradually.[4] Thus one should expect a negative time error because the sensation itself will fade. It is not necessary to appeal to a fading memory trace.[5]

In lifted weight experiments using the constant method, negative time errors are also extremely common (**6**, **7**, **21**). Again, however, it is not difficult to find a possible reason in the fading of the impression of the standard itself. In all standard experiments the subject lifts the weight and *puts it down*. In replacing it, the subject gains his last sensations from the weight, not when the proprioceptive impulses are at their maximum, but when the muscles are no longer sustaining completely the full pull of gravity on the weight. This implies that if the weight were dropped the negative time error would be eliminated. Such seems to be the case, according to preliminary results obtained in an experiment by McClelland and Sparks (**10**). Even if the negativity of the time error in such experiments cannot

[4] This simple physical fact may well account for the unexplained positive time errors obtained by Köhler (**8**) and Lauenstein (**9**) when the standard and comparison stimuli are separated by very short intervals of time (usually less than 2 sec.). "Die Bedeutung des positiven Zeitfehlers bei Kurzen Zwischenzeiten ist gegenwärtig noch vollkommen unklar" (**9**, p. 169). Over such a brief interval it is highly likely that the second stimulus falls on the receptor before it has entirely ceased responding to the first. Under such conditions it is hard to predict exactly what sort of interaction would occur between the two sound waves or between the two excitations of the hearing mechanism, but probably there would be an initial recruitment or reinforcement. Apparently the clicks are so brief and in such rapid succession that, in judging, the recruitment is assigned to the first stimulus and the second appears less. Such a hypothesis is highly conjectural, but it demonstrates the need for determining the interaction effect of the two sensory impressions, or the need for cutting off the after-effect of the first stimulus so abruptly that interaction is prevented. Before one of these conditions is fulfilled experimentally it seems premature to try and explain such a positive time error by psychological principles.

[5] Even when reverberation is controlled by the use of earphones as in the experiments of Needham (**15**), the negative time-error may still be explained by the 'off-effects' of the diaphragm in the earphone or of the ear itself. Stevens and Davis (**19**, p. 330) state, for instance: "When a stimulating tone stops abruptly, electrical activity does not immediately cease. Instead it subsides gradually. . . ."

be attributed to a fading stimulus source, it can be attributed to the reflex proprioceptive backlash which seems to accompany setting down the weight. The action potential records of Freeman and Sharp (**6**), Davis (**5**), and Payne and Davis (**16**) show this secondary input very clearly. Such a second, less intense, stimulation should act to depress the standard, just as an interpolated lighter weight did in the experiment by Guilford and Park (**7**). Once again the cause of the negative time error could well lie in the receptor adjustment, and need not be explained by a memory process at all.

Such considerations as these lay open once again the whole problem of the 'normal' negativity of the time error. It is quite possible that the whole theoretical structure which Pratt (**18**) has built on this fact, to show that memory traces disintegrate with time, will turn out to be irrelevant, since the phenomenon may be due to a fading stimulus source and may have nothing to do with memory. The fact that the negative time errors are so often found may mean no more than that nearly all sensations fade either because the physical stimulus fades or because receptors take a certain amount of time to stop responding.

Marchetti, in presenting lines tachistoscopically, happened to use stimuli which almost preclude entirely stimulus or receptor lag. This may easily explain why he failed to get the usual negative time error. The fact that his time errors were not only not negative, but positive can be explained by the assimilation factor which is also operative here.

Assimilation. (1) *Memorial.*—If the hypothesis stated in the introduction is correct—to wit, that Marchetti's positive time errors in judgments of visual extent were due to assimilation to longer contours in the background—then removal of those contours by presenting the lines in darkness should also remove the positive time error. The prediction is not entirely confirmed by the results of Condition J, however. In this condition, when the effect of a fading stimulus source is controlled, the time error is still positive, though not reliably so (C.R. = 1.27, $P < .11$). The result of Condition K which yielded no time error at all is further ground for believing that the positive time error obtained may be only a chance fluctuation.

If, however, the time error were reliably positive, it would not be difficult to explain. Even though the room was darkened the subjects carried with them in memory a total impression of many long contours from the room in which they were seated. It is easy to imagine that these long contours carried in memorial schemata would exert a sort of series effect (**15, 20**) on the memory trace of the standard. That is, the standard would be assimilated into the much longer memory series; it would therefore become longer, and the

second line would seem shorter by comparison. The only difficulty with this explanation is the question whether the room contours were sufficiently similar to the lines of light projected on the screen to interact with them at all.

Evidence that such interaction can occur is to be found in the data on absolute judgments of lengths of a line seen 30 seconds and 7 days previously. In this instance the older line has apparently been assimilated up to the other straight line contours in the every-day experience of the subjects, since it was a very short line relative to most of these lines. In this instance the influential experience is interpolated, rather than prior to the standard, but the effect is the same. The experiments of Guilford and Park (**7**), Turner (**20**) and Woodrow (**21**) show that assimilation may be proactive or retroactive.

The relevant data of this experiment, if they can be considered reliable enough to demand explanation, are then consistent with the interpretation of time errors as being sometimes due to assimilation of the standard up or down to memory 'standards' or schemata (series means), either present at the time the standard appears or formed between presentation of the standard and comparison stimuli. The memory standard in this case is built up largely by a series of impressions from a very large lecture room. It would be interesting to note what would happen if the experiments were performed in a small research room in which the lines could be much longer relative to the contours of the room. In this case, if the hypothesis is correct, the positive time error should lessen or disappear, provided, of course, that the sources of perceptual assimilation were also controlled.

(2) *Perceptual Assimilation.*—Under this heading fall the results which show that introducing frames around the lines to be compared decreases the negative time error. Particularly striking and significant is the difference between the time error in Conditions C and D. Here introduction of a square frame around the standard alone completely removed the negative time error caused by the fading stimulus source. Unquestionably the standard was assimilated up to the longer lines around it. This kind of assimilation has been called 'perceptual,' to contrast it with the influence of incidental, general, or series effects on the standard. That is, the stimuli which act upon the standard are present with it and not before or after. Any interaction between a standard and the background upon which it appears would be perceptual assimilation, though obviously the distinction between this and memory assimilation is purely formal and arbitrary. There is no evidence that the two kinds function differently.

Further evidence of such assimilation comes from a comparison of the combined mean of Conditions E, F and G with the combined

mean of A and B. The difference between these two means, though not completely reliable, supports the contention that introduction of frames, even though they be around both standard and comparison lines, causes an assimilation of the standard upward, as is evidenced by the shift of the time error in the positive direction.

So far the case for an assimilation hypothesis seems fairly complete, but there is one argument used in particular by Pratt (**18**), who has felt it to be inadequate, which should be answered. He points out that in experiments comparing the loudness of auditory intensities, introduction of a soft background sound will lower the standard more than will a phenomenally empty background which is presumably of zero or still less intensity. If the assimilation hypothesis were correct, he argues, the assimilation should be greater to zero than to a soft sound.

Such an argument is valid only if it can be assumed, as it certainly cannot, that a phenomenally empty ground is as similar to a sound as a soft sound is to the same sound. The probability is that the empty ground and the standard sound belong to two partly different trace systems, and hence interact with each other less than would two sounds. Pratt seems to have recognized the importance of the similarity factor earlier for he says: "For sounds the background is silence and silence seems not to stand in a close enough structural connection with the auditory impressions to serve as a source of assimilation" (**17**, p. 297). It seems therefore reasonable to assume that the contact between any two sensory impressions is determined by their similarity according to some function like the Skaggs-Robinson hypothesis (**11**, p. 462). If so, there should be less negative time error when the assimilation is to a phenomenally different ground than when it is to a similar sound.

Summary

The hypothesis was advanced that positive time errors in judgments of visual extent are due to assimilation between the lines presented and longer lines in the background. Groups varying in size from 11 to 39 subjects judged by the constant method the lengths of lines about a foot long under 10 different field conditions. The lines were presented to the subjects as a group by projection on a screen in a large lecture hall, and each subject made only 3 to 5 judgments per pair. The results are treated in terms of group means and their variability, a procedure which seems justified by analogy with experiments in other fields.

The conditions varied in three general ways: (1) In the presence or absence of room contours to which the lines could assimilate, (2) in the kind of frame in the immediate field of the lines, and (3) in the

XUM

control of the stimulus source. The results are expressed in terms of the points of subjective equality defined as the means of the distributions of equal judgments. Differences between these *PSE*s as determined by changes in the field conditions are evaluated according to the usual probability statistics. The following experimental conclusions seem justified:

(1) The presence or absence of large room contours has no reliable effect on the *PSE*. There is some indication that when the stimulus source is interrupted and the lines are presented in the dark, the time error will be positive.

(2) Introducing a square or rectangular frame around the standard stimulus or around both standard and comparison stimuli uniformly and reliably shifts the time errors in the positive direction.

(3) A fading stimulus source produces a reliable negative time error. Interruption of the stimulus source reliably shifts the time errors in the positive direction and eliminates the negative time error altogether.

(4) A line seen 7 days previously is judged to be larger (both wider and longer) than it is if seen 30 seconds previously.

These results led to the following theoretical conclusions:

(1) The negative time errors obtained almost universally in psychophysical experiments by the method of constant stimuli need not be explained by reference to a fading memory trace of the standard. In nearly every instance they may be attributed more parsimoniously to a fading in the sensation itself, caused either by a fading stimulus source or by receptor-adjustmants following stimulation.

(2) All remaining time error effects may be explained by assimilation, either memorial or perceptual, if it is further assumed that assimilation is subject to some such law of similarity as that which governs the amount of interaction between any two learning processes.

(Manuscript received May 4, 1943)

Bibliography

1. Boring, E. G. *Sensation and perception in the history of psychology.* New York: Appleton-Century, 1942, Chapter 15.
2. Burnham, R. W. Intersensory effects and their relation to memory theory. *Amer. J. Psychol.*, 1941, **54**, 473–489.
3. Culler, E. Studies in psychometric theory. *Psychol. Monogr.*, 1926, **35**, 56–137.
4. Culler, E. Studies in psychometric theory: XIV. On the probable error of the limen. *J. exp. Psychol.*, 1927, **10**, 463–477.
5. Davis, R. C. The pattern of muscular action in simple voluntary movement. *J. exp. Psychol.*, 1942, **31**, 347–366.
6. Freeman, G. L., & Sharp, L. H. Muscular action potentials and the time-error function in lifted weight judgments. *J. exp. Psychol.* 1941, **29**, 23–36.
7. Guilford, J. P., & Park, D. G. The effect of interpolated weights upon comparative judgments. *Amer. J. Psychol.*, 1931, **43**, 589–599.

8. KÖHLER, W. Zur Theorie des Sukzessivvergleichs und der Zeitfehler. *Psychol. Forsch.*, 1923, **4**, 115–175.
9. LAUENSTEIN, O. Ansatz zu einer physiologischen Theorie des Vergleichs und der Zeitfehler. *Psychol. Forsch.*, 1932, **17**, 130–177.
10. MCCLELLAND, D. C., & SPARKS, J. B. The effect of dropping the standard on the time-error in lifted weight judgments. (Unpublished paper)
11. MCGEOCH, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942, pp. 461–469.
12. MARCHETTI, P. Time-errors in judgments of visual extents. *J. exp. Psychol.*, 1942, **30**, 257–261.
13. MELTON, A. W. The methodology of experimental studies of human learning and retention. *Psychol. Bull.*, 1936, **33**, 305–394.
14. NEEDHAM, J. G. The time-error in comparison judgments. *Psychol. Bull.*, 1934, **31**, 229–243.
15. NEEDHAM, J. G. The effect of time interval upon the time-error at different intensive levels. *J. exp. Psychol.*, 1935, **18**, 530–543.
16. PAYNE, B., & DAVIS, R. C. The role of muscular tension in the comparison of lifted weights. *J. exp. Psychol.*, 1940, **27**, 227–242.
17. PRATT, C. C. The time-error in psychophysical judgments. *Amer. J. Psychol.*, 1933, **45**, 292–297.
18. PRATT, C. C. The law of disuse. *Psychol. Rev.*, 1936, **43**, 83–93.
19. STEVENS, S. S., & DAVIS, H. *Hearing, its psychology and physiology.* New York: Wiley, 1938.
20. TURNER, W. D. Intra-serial effects with lifted weights. *Amer. J. Psychol.*, 1931, **43**, 381–416.
21. WOODROW, H. Weight-discrimination with a varying standard. *Amer. J. Psychol.*, 1933, **45**, 381–416.
22. WOODWORTH, R. S. *Experimental psychology.* Chapters 17 and 18. New York: Holt, 1938.

BONE CONDUCTION DURING EXPERIMENTAL FIXATION OF THE STAPES [1]

BY KENDON R. SMITH

Princeton University

As is well known, sounds may be heard not only in the usual way, by air conduction, but also by transmission through the bones of the skull. When by accident or disease the passage of aerial sounds through the drum and ossicles is no longer possible, this second form of transmission engages special attention as the sole remaining route for hearing. The understanding of the process of excitation of the inner ear by bone conduction is therefore of great theoretical and practical importance.

The obvious explanation of bone-conduction hearing would be one of simple sound transmission. In this explanation, sound waves, spreading through the bones of the skull, are conceptualized as passing over into the fluid of the inner ear and thus stimulating the end organ. A related but more complex theory, the compression theory, holds that bone conduction is effective because the sound waves of compression and rarefaction which are produced spread practically instantaneously through the bones of the skull. Such waves would tend alternately to constrict and to enlarge the cavities of the skull. Thus, they would presumably bring varying pressures to bear upon the contents of the skull, among them the inner ear. A third theory, the inertia theory of bone conduction, maintains that there is an inertial lag of the various middle and inner ear structures behind the mass vibrations of the stimulated skull. The relative motions thus induced, between ossicular chain and skull, between endolymph and organ of Corti, and so on, are supposed to lead to stimulation of the inner ear.

A test of the inertia theory may be made by ascertaining the effect on bone conduction of a change in the effective inertia of the ossicular system. This may be done either by reducing or by increasing the effective inertia. The inertia may be reduced by severing the incudostapedial joint, and thereby removing the loading of the more peripheral parts: incus, malleus, drum, and the associated tendons and muscles. The inertia may be increased by any measure which adds to the mass of the moving system, or effectively restrains

[1] From the Psychological Laboratory of Princeton University.

its action. Perhaps the most complete restraint is afforded in the disease known as otosclerosis.

Otosclerotic deafness is caused by a disease of the bone in the region of the oval window. The bone first becomes spongiose, and then in healing it overgrows its normal boundaries with a scar which extends to the stapes and impairs its mobility. The condition is typically progressive, until at an advanced stage the stapes is effectively fixed, or ankylosed, in position, and is no longer capable of performing its normal function of transmitting vibrations to the fluid of the inner ear. The degree of fixation may reach practically 100 percent, as shown by reductions in acuity for air-borne sound of 60 decibels in cases of advanced otosclerosis. Although such air-conduction losses may be widespread in an irregular and unpredictable pattern throughout the audiogram, it is a striking fact that material losses in bone-conduction acuity are seldom reported in otosclerosis. (**1**, p. 7; **4**; **6**; **7**; **12**, p. 357; **13**, p. 290; **14**, p. 211.)

The present investigation attempted to simulate in the cat this condition of ankylosis, in order to determine whether or not stapedial motion relative to the skull is important in the responses of the ear to bone-conducted sound. Positive results would support the inertia theory of bone-conduction excitation and would also suggest that modification of stapedial motion might be involved in the etiology of some forms of bone-conduction deafness.

Procedure

Eight ears, from seven cats, were used in these experiments. Each animal was anesthetized by an intra-peritoneal injection of Dial liquid. It was then placed upon the operating table with the ventrum uppermost and the head held firmly in a jaw clamp.

The bulla was reached by a wide incision along the ventral midline of the neck. The shell and the septum of the bulla were extensively opened, the chorda tympani nerve and its cartilaginous support were removed from the bulla, and the mastoid bone ventral to the stapedius muscle and in the region of the facial nerve was partially cut away. A small brass screw was then turned into the posterior face of the mastoid bone, for eventual use in placement of the bone-conduction receiver.

To aid in the fixation of the stapes, a fine thread was passed around its neck. The thread was inserted between the stapedius tendon and the cochlea from the ventral aspect, passed around the dorsal side of the neck of the stapes, and returned out of the bullar incision. A few fine holes were drilled in the medial remnant of the shell of the bulla, and the two ends of the thread were looped through the holes with care to avoid any tension on the stapes.

Throughout the experiment, the animal lay in a sound-proof and electrically-shielded room. Stimuli were generated outside this room and led to the animal within, using the following set-up. The output of a General Radio Type 613-B beat frequency oscillator was filtered, adjustably attenuated, amplified, and delivered to a two-way, two-pole switch. With one setting of this switch, the current was made to activate a Western Electric Type 555 air transmission receiver. With the other setting it was sent through the primary winding of an impedance-matching transformer.

This much of the stimulating apparatus was outside the sound-proof room. Air-borne sound from the receiver was led into the room and to the operated ear through a tube whose inner end was firmly tied or taped into the external meatus. The current in the secondary winding of the transformer was carried by shielded cable into the room, and there made to activate a piezo-

electric crystal. The crystal was electrically and acoustically shielded, and equipped with a plunger which transmitted its vibrations. It was clamped during use in a heavy stand, which was adjusted to bring the plunger to bear upon the screw in the animal's mastoid bone. The plunger was directed in an anterio-dorso-medial direction, to reproduce the general direction of action of the usual bone-conduction stimulator. The air-conduction tones were of a high degree of purity, but the bone-conduction tones were slightly distorted under the circumstances.

The synchronous cochlear response was used to indicate inner ear excitation. To detect the response, a silver foil electrode was placed on the round window membrane, and an inactive needle electrode was placed in the skin or the superficial muscle of the incision. The responses so obtained were amplified and delivered to a General Radio Type 736-A wave analyzer for measurement.

With the ear under 1000~ stimulation by air conduction, an ordinary straight needle was passed between the two leads of the thread about the stapes. The point of the needle was firmly lodged in the bullar orifice of the Eustachian tube, and its shank was pulled tightly against the reduced mastoid by a tie to the screw in that bone. The response was observed during this manipulation, and the needle was brought as close to the neck of the stapes as possible without causing a material change in the response. This placement of the needle aided in the eventual fixation of the stapes.

The animal was curarized by an injection into the femoral vein, in order to suppress the activity of the aural muscles. Since the necessary degree of curarization also depressed respiration, the animal was connected by a tracheal canula to an artificial respirator which functioned throughout the procedures described below.

The ear was stimulated by air and by bone conduction, first in the normal condition, then with the stapes in fixation, and finally with the stapes released and the aural apparatus returned as nearly as possible to normal.

In the initial normal condition, separate air- and bone-conduction tones in the frequency range of 100~ to 10,000~ were chosen as stimuli. These tones were of an intensity just sufficient to produce a response of a standard magnitude, varying in different series of measurements from 1 to 110 microvolts.

Following these normal determinations, tension was exerted upon the thread around the stapes, and the thread was then fastened to the shell of the bulla. The neck of the stapes was pulled securely against the body of the needle, and stapedial motion relative to the skull was thereby restricted. The degree of fixation produced by this method was judged during the procedure itself by observing the diminution in response to a 1000~ air-conduction tone. In different fixations, it varied from 12 to 55 decibels, or, in linear terms, from 75 to 99.8 percent. After this fixation, the previous testing procedure was repeated.

Finally, the tension was removed from the stapes, and measurements were again repeated. This second group of normal measurements, when compared with the initial normal measurements, showed whether the fixation process had caused any irreversible damage to the transmission system.

This sequence of initial normal, fixational, and second normal tests was often repeated on a given ear, so that in all there were 15 such series of tests.

Results

Fig. 1 shows an illustrative set of results for a single series. In this figure, the abscissa represents the frequency of the tones used as stimuli: in the separate case of either mode of conduction, the stimulus used to obtain the initial normal response at any given frequency was also used to obtain the fixational and the second normal response at that frequency, with the result that all responses at the frequency were obtained with the same intensity of tone. The ordinate shows the change in response from the corresponding standard response, in decibels, for each condition of stimulation. Air-conduction readings are indicated by *x*'s, and bone-conduction by dots, while solid-line curves represent the fixational responses, and broken-line curves

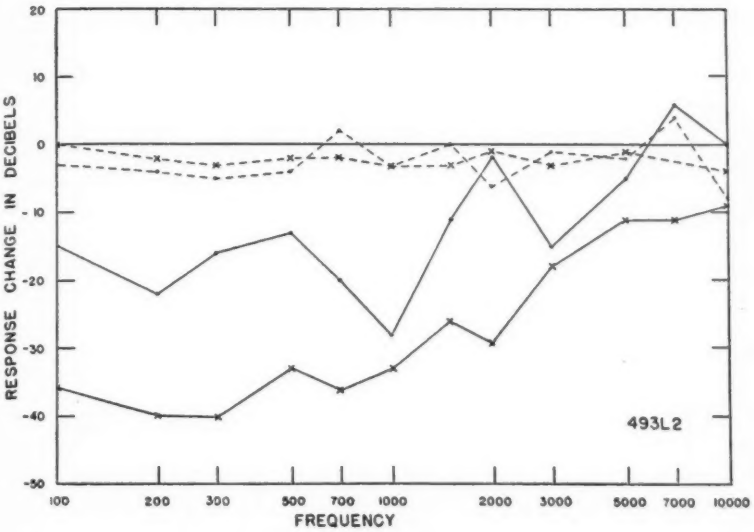


FIG. 1. Change in response with stapedial fixation (solid-line curves) and release (broken-line curves), by air conduction (*x*'s) and by bone conduction (dots). Second series, left ear, cat 493.

the second normal. The zero line corresponds to the standard, or initial normal, response readings, which in this case were all of a magnitude of 10 microvolts.

As may be seen from Fig. 1, stapedial fixation rather profoundly depressed air-conduction responses in this series. The depression, though general, was less marked in the higher frequencies. For the lower range of tones, below 1000~, the decrease in response due to stapedial fixation exceeds 30 decibels and in some cases becomes as much as 40 decibels. The effect is progressively smaller as the frequency increases, until at 5000~ and above it is 11 decibels or less. The air-conduction responses obtained after release of the tension show only slight differences from their respective standard responses.

The bone-conduction responses shown in Fig. 1 fall into a somewhat different pattern. Here stapedial fixation also led to a substantial depreciation in response, but the effect was most pronounced in the middle frequencies. Lower frequencies, in the range of 100~ to 500~, showed some loss, but not as much as the 28 decibel loss at 1000~. The changes at and above 5000~ were again small, being in all cases less than 7 decibels. The agreement between second normal and initial normal responses is somewhat less than for the air-conduction form of stimulation, but is still satisfactory; the slight

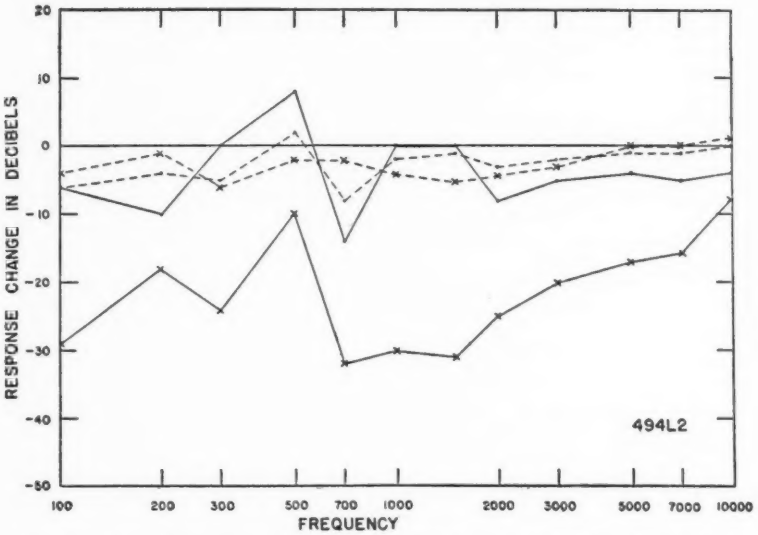


FIG. 2. Change in response with stapedial fixation (solid-line curves) and release (broken-line curves), by air conduction (x's) and by bone conduction (dots). Second series, left ear, cat 494.

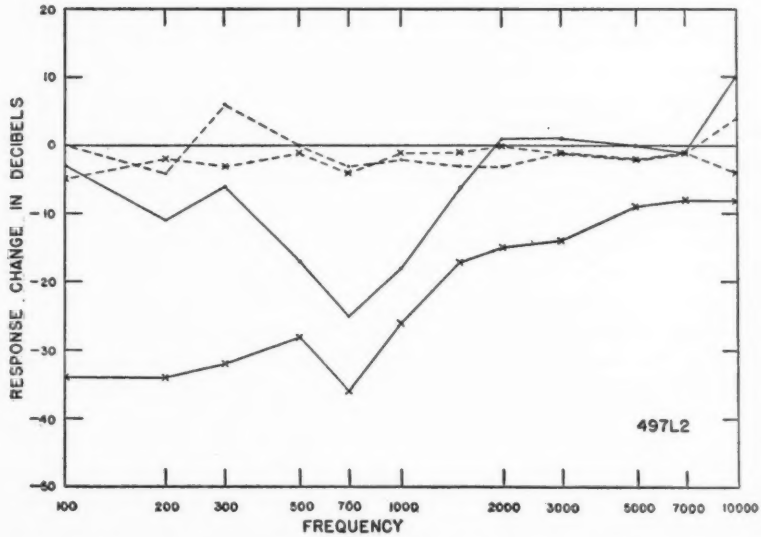


FIG. 3. Change in response with stapedial fixation (solid-line curves) and release (broken-line curves), by air conduction (x's) and by bone conduction (dots). Second series, left ear, cat 497.

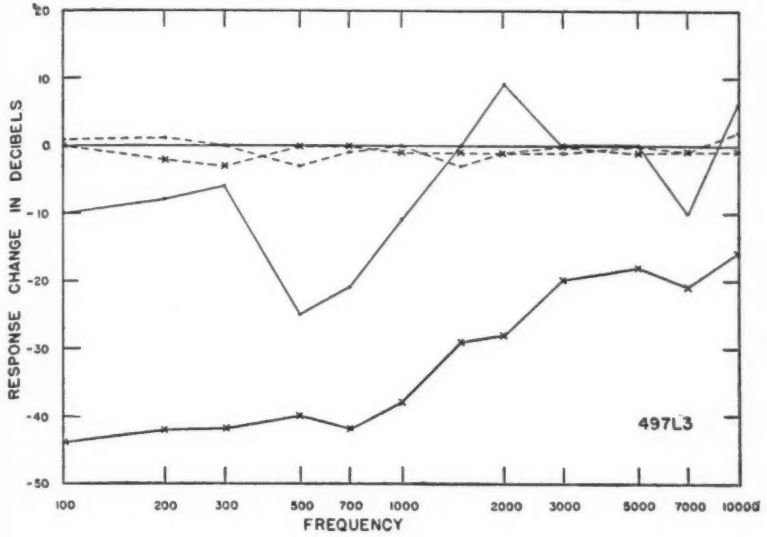


FIG. 4. Change in response with stapedial fixation (solid-line curves) and release (broken-line curves), by air conduction (*x*'s) and by bone conduction (dots). Third series, left ear, cat 497.

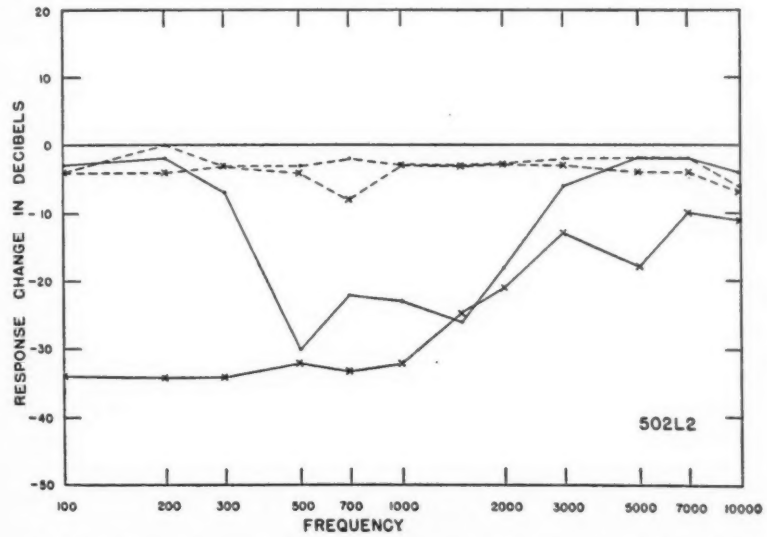


FIG. 5. Change in response with stapedial fixation (solid-line curves) and release (broken-line curves), by air conduction (*x*'s) and by bone conduction (dots). Second series, left ear, cat 502.

differences no doubt arise from inherent instability in the bone-stimulation system.

The same general pattern which appears in Fig. 1 is repeated in Figs. 2, 3, 4, and 5. With the same conventions as the first figure, these graphs show the records of four additional series of standard, fixational, and second normal tests. The profound losses in the lower frequency range, the less severe losses as the frequency rises above 1000~, and the minor deficiencies in the higher range of frequencies, are all again evident in the case of air-conduction responses. Bone-conduction responses show depression by stapedial fixation mainly in the middle frequencies in these four series as in the one shown in Fig. 1. The lower frequencies are sometimes affected to a lesser degree in the same direction, but the higher frequencies remain practically normal. These results are least prominent in Fig. 2, but still discernible.

A comparison of the results by air and bone conduction is most easily made in Fig. 6, which is a composite of all of the first five graphs. The arrangement of the graph is the same as above. The losses at lower and middle frequencies in the air-conduction curves, and the losses at middle frequencies for bone conduction, are readily apparent.

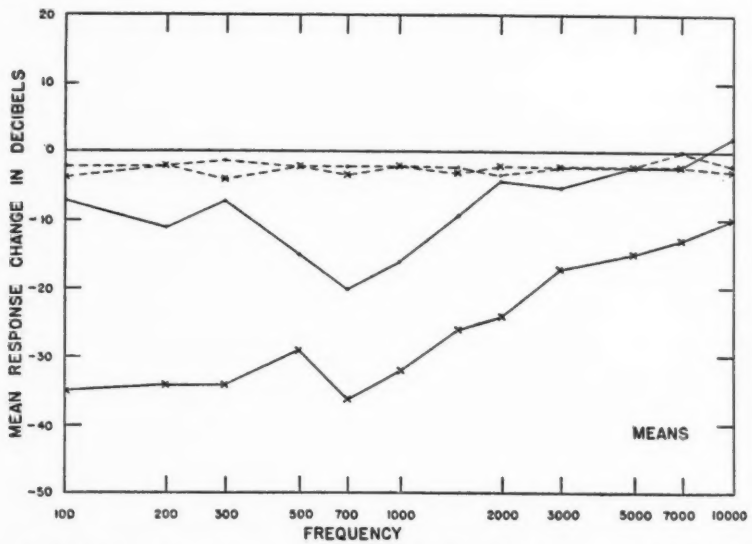


FIG. 6. Mean change in response with stapedial fixation (solid-line curves) and release (broken-line curves), by air conduction (x's) and by bone conduction (dots). Composite of first five figures.

As far as the relationship between standard and fixational responses is concerned, the graphs shown are typical of what is found in all tests. These records were, however, selected from the series in which the second normal response agreed with the appropriate initial normal response within a range of plus or minus 8 decibels. In the remaining trials there was less agreement between initial and second normal responses. In these tests, as nearly as could be determined from observation at the time, the discrepancies between standard and second normal responses were frequently due to factors which did not vitiate comparison between standard and fixational responses. After release of fixation, the thread often remained in light contact with the stapes, weighting it in a fashion different from that during initial normal tests. Also, some fixations were probably drastic enough to produce slight misalignments of the aural apparatus. These would probably not be serious as regards the results obtained during fixation, but would show up afterwards in modifications of the second normal responses.

These results show, then, that fixation of the stapes relative to the skull, when sufficient to reduce profoundly the synchronous cochlear response to air conduction, also substantially reduces the responses to bone conduction in the middle frequencies. Stapedial motion relative to the skull is thus important in the bone-conduction responses of the inner ear in the middle frequency range.

Discussion

The results given above are of course limited to the cat; but it is clear that within limits the inertia theory is substantiated, though without necessary refutation of the other theories mentioned. For middle frequencies, responses by bone conduction are materially affected by stapedial fixation. We must now inquire why inertia stimulation should be important primarily or solely in the middle frequency range.

Significant for this problem is evidence from another direction of investigation. Experiments with practically instantaneous stimuli have indicated that the resonant frequency of the cat's aural system is probably within this middle range of from 500~ to 1500~ (**5**; **13**, p. 262). The immediate suggestion is that inertial stimulation becomes important only as the stimulus frequency approaches the ear's resonant frequency and thus sets it in motion with greater and greater ease. Such a hypothesis would explain the great importance of inertial stimulation in the central frequency range, and its waning importance toward the ends of the range, and would predict the loss patterns found in the present investigation.

This hypothesis may also serve to reduce the apparent conflict between the results of this experiment and the usual report that bone-conduction deafness of material degree is seldom found in otosclerosis. If inertial stimulation and stapedial fixation are important only in the resonant region of the ear, bone-conduction losses in otosclerosis would be expected mainly in the middle frequencies, the resonant frequency of the human ear apparently being approximately that of the cat's ear, or slightly higher (**1**, p. 203; **3**; **13**, p. 262). The common tests at only relatively low frequencies in cases of otosclerosis would thus neglect any regions of substantial bone-conduction losses, and general normality might be assumed.

A supplementary explanation of the difference between clinical and experimental report lies in the possibility of a limit to the effects of stapedial fixation. It may be that fixation in the human subject never produces bone-conduction deafness beyond a moderate degree, and that other mechanisms, probably the ones responsible for bone-conduction hearing outside of the middle frequency range, then maintain function. Mechanisms of this sort may readily be imagined in the inertial action of the contents of the inner ear, in compression effects, and in the direct transmission of sound from skull to inner ear fluid. Any or all of these might prevent a very marked loss of bone-conduction hearing upon stapedial fixation. If the limit of loss should place acuity near the lower end of the range of acuity considered normal clinically for a given tone, depreciation down to that limit might yield diagnosis of normality, and the real loss go undetected. The existence of such a limited effect is suggested by the experimental results of Békésy (**2**) and of Bárány (**1**).

Both experimenters attempted to fix the human stapes relative to the skull. Békésy's (**2**) method was to raise the air pressure in the open external meatus, and thus to force the ossicular system into an extreme and rigid position. The pressure change was accomplished by the use of a large vessel which made air-tight junction with the head about the external ear. The experimental condition was found to produce depreciation in both air- and bone-conduction acuity under proper circumstances. Air conduction suffered markedly more than bone conduction. This observation showed almost certainly the action in bone conduction of a mechanism supplementary to the ossicular chain: if the ossicles were the sole mechanism of bone conduction, their fixation should affect bone-conduction as much as it did air-conduction hearing.

Bárány (**1**) accomplished the same sort of fixation by evacuating air from the tympanic cavity. The vacuum was produced by inspiration with closed nostrils and mouth, and was relieved gradually by careful admission of air through the Eustachian tube. His re-

XUM

sults were similar to those described for Békésy's experiment, and showed in addition an apparent lower limit of bone-conduction loss. An increase in the tympanic vacuum beyond the point at which bone-conduction loss was about 12 decibels apparently did not produce further bone-conduction deafness, although air-conduction loss increased beyond this critical point. As in Békésy's experiment, there may be some question here as to the efficacy of stapedial fixation and also as to the effects of middle ear pressures upon the action of the inner ear (see Hughson and Crowe, **8** and **9**).

In the same 1938 report, Bárány described a study of the interaction of various mechanisms in bone conduction which is of theoretical interest. He extended Békésy's (**2**) preliminary work, which showed that a bone-conduction tone could be cancelled subjectively by an air-conduction tone of the same frequency and of proper amplitude and relative phase.

Bárány attempted to find the comparative importance of the inertia and of the compression components in inner ear stimulation by bone conduction. He reasoned that compression stimulation of the inner ear would remain in the same phase relationship with the bone-conduction receiver, independently of the receiver's position or direction against the skull, because of the practically instantaneous spread of sound waves in bone. The phase relationship of the inertia stimulation to the receiver would, however, be changed by as much as 180° by changing placement of the receiver on the head. The amount of change in the phase relationship between the resultant stimulation of the inner ear and the receiver, as position and direction of stimulation were changed, was thus taken as an index of the comparative importance of the inertial and of the compression effects.

Bárány inferred the nature of this relationship from the phase relationship between the air-conduction tone required for cancellation of the bone-conduction tone, and the bone-conduction receiver. He found considerable change in phase of inner ear stimulation by bone conduction relative to phase of receiver, amounting in some cases to nearly 180°, as the receiver was moved about the head. Bárány concluded that bone-conduction stimulation depends largely upon inertial effects. However, since his bone-conduction tones were accompanied, perhaps unavoidably, by interfering air-conduction tones whose influence was questionably evaluated, his findings are apparently tentative.

A number of animal experiments, as well as these experiments of Bárány and Békésy, have indicated that bone-conduction audition depends upon the state of the middle ear, but the evidence has in some cases been inconsistent. Hughson, Thompson, and Witting (**10**) found that lesions of the bullar septum, the incudostapedial

joint, or the tympanic membrane tended to depress VIIIth nerve responses to bone conduction; tension on the tensor tympani tendon also impaired bone-conduction responses, but these responses were enhanced by the mere opening of the bulla. Wever and Bray (**15**) found varying losses, none greater than 20 decibels, in the cochlear response to bone conduction upon luxation of the incudostapedial joint, and Guild (**7**) reported no significant change in such responses when the posterior incudal ligament was detached from the incus. Kobrak (**11**) used the reflex response of the middle ear muscles as a criterion of audition, and concluded that modification of the middle ear always leads to increased, rather than decreased, bone-conduction acuity.

The outcome of the present study joins the results of some of these animal experiments in emphasizing the possibility of human bone-conduction deafness arising from middle ear pathology. Guild (**7**) has already demonstrated a high correlation between clinical diagnosis of material bone-conduction deafness and an abnormality of the skull structure in the mastoid region. As shown by post-mortem histological preparations, such diagnoses were made in all cases in which later microscopic study revealed tiny fractures of all the bony trabeculæ in this region, and in no cases with some intact trabeculæ. Together, these lines of evidence throw doubt upon the traditional practice of attributing bone-conduction deafness to neural deficiency.

Summary and Conclusions

In experiments on the cat, stapedial motion relative to the skull was restricted in 15 series of tests on 8 ears. It was found that fixation of the stapes relative to the skull, when sufficient to reduce profoundly the synchronous cochlear responses to air conduction, also substantially reduces responses to bone conduction in the middle frequencies. Stapedial motion relative to the skull is thus important in the bone-conduction responses of the cat's inner ear to tones of middle frequency.

The investigation indicates the clinical importance of middle ear lesions in the etiology of bone-conduction deafness. The results also lend partial support to the inertia theory of bone conduction. An explanation of the fact that the ossicular system is important for bone conduction only within the middle range of frequencies is suggested in terms of the presence and influence of the resonant frequency of the ear.[2]

[2] The writer wishes to acknowledge his debt and appreciation to Professor E. G. Wever for his aid in the design and execution of this experiment, and to Mr. David Van Iderstine for his valuable technical assistance.

References

1. Bárány, E. A contribution to the physiology of bone conduction. *Acta Oto-Laryngol.*, 1938 (Supplement), **26**, pp. 227.
2. Békésy, G. v. Zur Theorie des Hoerens bei der Schallaufnahme durch Knochenleitung. *Ann. d. Physik*, 1932, **13**, 111–136.
3. Békésy, G. v. Ueber den Knall und die Theorie des Hoerens. *Physik. Z.*, 1933, **34**, 577–582.
4. Bunch, C. C. Some observations on clinical otosclerosis. *Ann. Otol.*, 1934, **43**, 344–366.
5. Davis, H., Derbyshire, A. J., Lurie, M. H., & Saul, L. J. The electric responses of the cochlea. *Amer. J. Physiol.*, 1934, **107**, 311–332.
6. Guild, S. R. Early stages of otosclerosis. *Arch. Otolaryng.*, 1930, **12**, 457–483.
7. Guild, S. R. Hearing by bone conduction: the pathways of transmission of sound. *Ann. Otol.*, 1936, **45**, 736–754.
8. Hughson, W., & Crowe, S. J. Function of the round window. *J. Amer. med. Ass.*, 1931, **96**, 2027–2028.
9. Hughson, W., & Crowe, S. J. Immobilization of the round window membrane. *Ann. Otol.*, 1932, **41**, 322–336.
10. Hughson, W., Thompson, E., & Witting, E. G. An experimental study of bone conduction. *Ann. Otol.*, 1936, **45**, 844–858.
11. Kobrak, H. Utilization of the tensor tympani muscle reflex in the analysis of the function of the middle and internal ear structures. *Ann. Otol.*, 1936, **45**, 830–837.
12. Politzer, A. *Diseases of the ear.* London: Baillière, Tindall, & Cox, 1909.
13. Stevens, S. S., & Davis, H. *Hearing.* New York: John Wiley & Sons, 1938.
14. Walkyn-Thomas, F. W., & Yates, A. L. *Principles and practise of otology.* Baltimore: William Wood & Co., 1933.
15. Wever, E. G., & Bray, C. W. The nature of bone conduction as shown in the electrical responses of the cochlea. *Ann. Otol.*, 1936, **45**, 822–830.

THE RELATIVE DIFFICULTY OF MORSE CODE ALPHABET CHARACTERS LEARNED BY THE WHOLE METHOD

BY S. D. S. SPRAGG

Queens College

Introduction

Learning to recognize the International Morse code alphabet characters through auditory presentation is a 'paired-associates' learning situation of considerable interest for the psychology of learning, theoretical as well as practical. The incorrect responses of *S*s learning to receive code can readily be described with the terminologies which have been employed in the more traditional perceptual and learning situations. For example, when an *S* perceptually confuses two alphabet characters, the difference between them can be described as being below the *S*'s 'threshold of discrimination,' under the conditions of the moment. Or, we might use the terminology of the conditioned response and speak of an *S*'s making a 'generalization response' from one stimulus character to a similar one.

Most of the current methods of primary code instruction employ some variation of the part method of learning. The alphabet is usually divided into four or five groups, sometimes on an arbitrary basis, but more often on the basis of alleged differences in difficulty. It is true that inspection reveals a few of the alphabet characters to be obviously easy, and some others to be just as obviouslydifficult; but there are many that cannot readily be assayed *a priori* as to their difficulty simply by inspection of their elements and pattern.[1] A high degree of confusability with another alphabet character may make a given character difficult to learn, out of all proportion to its intrinsic complexity. In a training situation the difficulty of an alphabet character may be significantly altered depending upon whether it is presented in a group containing other characters which have either high or low confusability with it.

Despite the fact that code has been taught for years there is little or no *experimental* evidence as to the relative difficulty of the code characters. Such evidence is needed in order to achieve effective grouping of alphabet characters in primary code instruction. It can be most directly obtained by analysis of data from *S*s trained by

[1] See Table I, columns 2 and 3.

the whole method. This does not intend to imply that the whole method is necessarily superior in the practical training situation, but it is necessary to employ the whole method of training if it is desired to compare accurately the relative difficulty and confusability of the alphabet characters.

The difficulties and confusions encountered in mastering code also make available to the experimental psychologist a fairly difficult learning situation which should be a fruitful field for the investigation of a wide variety of learning and perceptual principles.

Problem

The present experiment is an attempt to determine (*a*) the relative difficulty of International Morse code alphabet characters, and (*b*) the kind and amount of confusion associated with each character. Some observations were also made on the rate of learning.

Method

Ss were undergraduate college students of both sexes, members of a non-credit code class for which they had formally registered and paid a $1 laboratory fee. Motivation was very good.

The code characters were presented by means of a Gardiner-Levering tape sender, audio-oscillator, and small loud-speaker. The dots and dashes cut in the tape were spaced in such a manner that each character was sent at a tempo corresponding to a sending rate of 13–15 words per min. This procedure facilitates the *Ss* learning each character as a pattern rather than analyzing it into its elements. The complete whole method of training was used. After each alphabet character had been demonstrated by naming it and sending it twice, the *Ss* were thereafter trained on formal lessons. Each lesson contained 100 alphabet characters arranged in random order. Several lesson tapes were cut, and used in rotation; thus all letters of the alphabet appeared with very close to equal frequency. Three lessons were presented at each class period. The class met three times a week.

The very effective procedure for primary code training developed by Keller (1) was followed here. After the code character sounded, the *S* was allowed three sec. in which to write down the letter for it, then the instructor spoke the phonetic equivalent of the letter. *Ss* were required to correct every omission and incorrect response by writing the correct letter just below the response space, as soon as the instructor named the character. The prepared lesson sheets were of the type developed by Keller.

When most of the class had mastered the code alphabet in this manner, they were shifted to receiving at a five words per min. rate, then at increased rates as their progress warranted. The present experiment is concerned with performance during the primary training. For this purpose analysis was made of the lesson records of all *Ss* who reached a criterion of five or fewer errors (in a lesson of 100 items) before the class was shifted to the five w.p.m. rate. This was primarily a code class, and it was not felt justified to hold back the group until everyone had reached the criterion set. Thus, the data are drawn from the top 65–70 percent of the code class and represent a selected group—selected in the sense that it consists of those persons who were able to meet the criterion in the time allowed.

The group thus emerging consisted of 19 *Ss* (14 men, 5 women). For each *S* an analysis was made of the frequency and kind of errors made on each lesson, from the initial one up to and including the first lesson on which the criterion was met. Some 8895 errors were thus tallied and classified.[2]

[2] Mr. Benson Bieley assisted in the tabulation of some of these data.

Results

The *Ss*, all inexperienced in receiving code at the beginning of training, required from 7 to 27 trials to meet the criterion. Only two *Ss* required more than 18 trials, and the mean of the group was 14 trials (median = 12 trials).

A graphic summary of the rate of learning for the group was secured by constructing a Vincent type curve of error elimination. Fig. 1 shows the mean number of errors at successive fifths of learning.

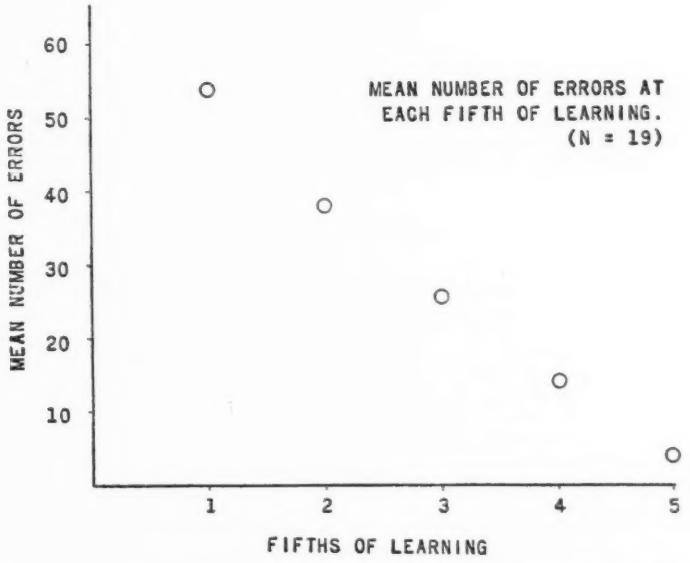


Fig. 1. Showing the mean number of errors made at successive fifths of learning Morse code alphabet characters. (N = 19.)

It will be noted that the improvement, when plotted in this manner, is almost linear.

Detailed results of the experiment are summarized in Table I, in which are presented the rank order of difficulty of each alphabet character, the total number of errors made on each, the total number of omissions and substitutions, and the two most frequently substituted letters for each stimulus letter, together with the percent that each is of the total number of substitutions made on that letter.

An examination of column 4 of the table reveals at once that the alphabet characters show clusters or 'families' of difficulty. First we have the three most difficult characters—W, P, and Y. Next

TABLE I

SHOWING THE RANK ORDER OF DIFFICULTY OF MORSE CODE ALPHABET CHARACTERS, THE TOTAL FREQUENCY OF ERRORS, THE FREQUENCY OF OMITTED RESPONSES, THE FREQUENCY OF SUBSTITUTIONS, THE TWO MOST FREQUENTLY SUBSTITUTED LETTERS, AND THE PERCENT THAT EACH IS OF THE TOTAL NUMBER OF SUBSTITUTIONS FOR THAT LETTER. (N = 19)

Order of Difficulty	Stimulus Letter	Code Character	Total Errors	Total Omissions	Total Substitutions	Most Frequently Substituted Letters			
						1st	%	2nd	%
1	W	·— —	666	363	303	R	33	U	17
2	P	·— —·	643	321	322	J	20	L	16
3	Y	—·— —	612	289	323	C	24	L	15
4	X	—··—	549	243	306	B	38	P	10
5	Q	— —·—	509	223	286	Z	36	L	18
6	J	·— — —	507	264	243	P	18	L	14
7	F	··—·	498	260	238	L	19	Y	14
8	L	·—··	489	264	225	Z	21	P	13
9	U	··—	480	216	264	S	41	O	14
10	G	— —·	463	243	220	W	20	D	13
11	C	—·—·	445	216	229	Y	31	P	11
12	K	—·—	429	233	196	R	24	D	15
13	Z	— —··	413	203	210	Q	43	L	13
14	D	—··	380	217	163	K	17	G	14
15	R	·—·	370	191	179	W	25	K	21
16	B	—···	295	143	152	X	21	Y	9
17.5	A	·—	183	60	123	I	34	N	33
17.5	V	···—	183	81	102	H	51	U	9
19	H	····	159	32	127	S	47	V	28
20	N	—·	135	50	85	A	55	M	25
21	M	— —	126	45	81	N	38	I	21
22	I	··	116	28	88	A	47	N	32
23	S	···	97	45	52	O	29	H	27
24	O	— — —	69	37	32	S	22	M	16
25	T	—	57	14	43	E	74	—	—*
26	E	·	22	9	13	T	38	—	—*

* Scattered.

follows a family of 12 characters of considerable, but fairly comparable, difficulty. Then comes a transitional character—B. Finally we have a family of 10 relatively easy characters. The break between the two major clusters is a sharp one. A, which ranks 17.5 in order of difficulty, is less than half as difficult as R, which ranks 15.

As for results of individual *S*s, W ranked first in order of difficulty for 7 *S*s, P and Y for 3 each, Q and F for 2 each, and L and U for 1 *S* each. W, P, and Y ranked very high in order of difficulty for every *S*.

That the rank order of difficulty found here has a high degree of reliability is indicated by two lines of evidence. First, the 19 *S*s in this experiment were drawn from a Fall 1942 group and a Spring 1943 group (of 9 and 10 *S*s respectively). The rank-difference correlation coefficient between the orders of difficulty for these two sub-groups is $+.94 \pm .02$. Second, in a recent experiment of Keller and Taubman (2), alphabet and digit characters were taught together

by the whole method. Ranking in order of difficulty the alphabet characters only, there is a rank-difference correlation coefficient of $+.91 \pm .02$ between Keller and Taubman's order of difficulty and that reported here. It would seem that the relative order of difficulty of the alphabet code characters has been well established.

Errors in receiving Morse code are made in two ways: responding with an incorrect letter (a substitution response), or failing to make a response (an omission). Of the 8895 errors made by the group, 52 percent (4605) are substitution errors; the remainder are omissions. This statement conceals two facts, however. First, errors of omission are relatively more frequent during the early part of training, less so as the *Ss* approach the criterion. Second, *Ss* tend to make relatively more errors of omission on the difficult characters, and more errors of substitution on the easy characters. If we group the 15 most difficult characters (which takes us to the transition point described above) we find a mean of 50 percent substitution responses. The group comprising the 11 easiest characters has a mean of 63 percent substitution responses. The difference between the two means is statistically reliable, being 4.4 times its standard error.

A Morse code character may be difficult either because of its complexity or because it is confused with another character. Of the eight most difficult characters all but one are four-element characters (four is the maximum number of elements found in the Morse code alphabet). It is interesting to note however, in view of the role of confusability to be discussed below, that the most difficult character in the alphabet—W—is a three-element character.

Errors of substitution indicate, of course, a confusion between two or more characters. A tally was made of the frequencies of all substitutions for each alphabet character. In the last four columns of Table I is presented a summary of the two most frequently substituted letters for each stimulus character, with the frequency expressed as a percent of the total number of substitutions for each letter. For example, 303 substitution errors were made on the letter W; of this number 33 percent were substitutions of R, 17 percent were substitutions of U.

Substitution errors can be classified into several categories in terms of the kind of resemblance shown between the stimulus character and the character for the substituted response. For 16 of the alphabet characters the most frequently substituted letter is one whose code character has the same number of elements as the stimulus character and differs from it only in the nature of its final element. For example, Y (dah-dih-dah-dah) is confused most frequently with C (dah-dih-dah-dih). The major confusions of the six most difficult alphabet characters are all in this category (and of 14 of the first 18

characters in order of difficulty). Of the eight easiest characters only two have major confusions in this category. Confusion of all-dot with all-dash characters, or vice versa, is the most frequent substitution for four characters; reversal of order of elements accounts for three characters, and miscellaneous reasons for the other three characters.

This analysis reveals clearly that the greatest single source of confusion in learning to discriminate between Morse code characters, especially the more difficult ones, lies in confusing the character with another one which is identical with it except for the final element in its pattern.

Discussion

This experiment has not attempted to deal directly with the important practical problem of the most effective grouping of the code alphabet characters in order to produce highest efficiency in learning. The results of the present experiment have, however, demonstrated with a high degree of reliability the relative order of difficulty of the alphabet characters in primary code learning. It has also been possible to show, both in specific instance and in general principle, the most important kinds of confusion between the alphabet characters.

On the basis of the results reported here it is possible to construct groups of alphabet characters of almost any desired relationships in difficulty and of widely varying amounts of confusability within each group. For example, a set of four groups of approximately equal difficulty and with very low confusability within groups would be the following:

Group I: W, J, L, D, H, T
Group II: P, U, Z, R, V, I
Group III: Y, G, K, A, B, S, M
Group IV: X, F, Q, C, N, O, E

Or, if it should be desired to test the efficiency of learning code when characters of maximum confusability are grouped together, an analysis of the results of the present experiment suggests this arrangement:

Group I: B, X, C, Y, Z, Q
Group II: D, K, R, W, G, O
Group III: F, L, H, S, V, U
Group IV: P, J, M, A, I, N, T, E

This arrangement of groups has a descending order of difficulty and contains a great deal, perhaps a maximum amount, of intra-group

confusability. Ss who learn code according to this grouping will have differentiated all the difficult discriminations by the time they have mastered the separate groups, and further differentiations when the groups are combined should be relatively easy. Many other methods of combining the alphabet into groups can be worked out on the basis of the present results.

Answers to such questions as whether code can be learned most efficiently by making the groups of equal difficulty, by teaching the most difficult characters first, or last, etc., are research problems of pressing practical importance. The present study, and also that of Keller and Taubman (**2**) are basic contributions to these further problems in that they supply for the first time definite *experimental* evidence concerning the difficulty and confusability of the International Morse code alphabet characters.[3]

Summary

Analysis was made of the errors made by 19 Ss in primary training in receiving International Morse code alphabet characters. A highly reliable rank order of difficulty for the alphabet characters was established. The kind and amount of confusion between characters was determined, and it was shown that maximum confusions occur between alphabet characters that are identical with each other except for their final element.

(Manuscript received May 25, 1943)

References

1. Keller, F. S. Studies in International Morse code. I. A new method of teaching code reception. *J. appl. Psychol.* (To appear)
2. Keller, F. S., & Taubman, R. E. Studies in International Morse code. II. Errors made in code reception. *J. appl. Psychol.* (To appear)

[3] Since this paper was written the survey of Taylor (*Psychol. Bull.*, 1943, **40**, 461–487) has appeared. His account corroborates (p. 466) the present writer's assertion of a paucity of reliable experimental evidence concerning the relative difficulty of the code alphabet characters.

AN EXPERIMENTAL COMPARISON OF CODE-LEARNING METHODS

BY JOHN P. SEWARD

Connecticut College

PROBLEM

The present study was instigated by the work of Dr. Fred S. Keller, of the Subcommittee on Learning and Training of the Emergency Committee in Psychology. His work was aimed at improving methods of teaching International Morse code reception in training camps. In spite of the success of his efforts there appeared to be a real need for a controlled comparison of various procedures, a need which Dr. Keller himself was the first to emphasize. The same need was recently pointed out by Dr. Trabue in the report of the subcommittee (**10**). In addition to its practical desirability, such a study should throw further light on the memorizing process. For a survey of the literature on learning telegraphic code the reader is referred to Taylor's recent review (**7**).

The experiments here reported cover only a small segment of the problem. In the first place, they are limited to the initial stages of learning to receive code. The small number of subjects available made it necessary to use each subject as his own control. Thus, a subject could learn only part of the alphabet by any one method. The limited time which subjects could spare made it impossible to prolong the experiments until high speeds of reception had been attained.

In the second place, blinker signals were employed throughout. This choice was dictated by local considerations. Whether the findings are applicable to radio communication is a question which will be considered below.

In the third place, the subjects, undergraduates at a women's college, represent a rather narrow sampling of the population. The wide range of individual differences shown in code learning, however, suggests that the basis of selection has little to do with this ability.

Following is a general description of the learning methods compared:

Dot-dash Method (*Method D*)*:* This method represented an attempt to approximate under controlled conditions a conventional procedure in learning the Morse alphabet. Subjects were given a fixed time of several minutes in which to study a printed list of letters with their signals in dots and dashes. They were then given one or more blinker tests in which the signals were flashed in random order, after which they heard the letters read off and corrected their mistakes. The cycle was then repeated.

Prompting Method (Method P): This method incorporated the essentials of the one developed by Keller (2) to replace older procedures. Two basic principles are involved in Keller's method: (1) The signals are presented from the very beginning in the form in which they are finally received (a principle already advocated by Koch (3, 5)). Dot-dash equivalents are considered superfluous, if not positively harmful. (2) In the early stages of training, each attempted response to a signal is followed by announcement of the letter, thereby providing 'reinforcement' of a correct response and revision of a wrong one. In the present version, each letter was first named and signaled. The signals were then flashed in random order, separated by brief intervals in which the instructor named the letter and the subjects confirmed or corrected their own responses. This procedure continued for a prescribed period, culminating in a test.

Combined Method (Method C): This method combined the two just described. After several minutes in which the subjects studied printed lists of dot-dash equivalents, the instructor named and signaled the letters, then flashed and named them in random order as in the prompting method, and finally concluded the cycle by a test. It was hoped that this method would combine the advantages of the other two.

General Procedure

All of the experiments followed a common plan. When two methods were compared, the subjects were divided into two groups and two lists of letters of approximately equal difficulty were selected from the alphabet. The methods were then rotated by the conventional procedure, as illustrated below:

	List I	List II
Group A	Method X	Method Y
Group B	Method Y	Method X

When three methods were compared they were similarly rotated among three lists and three groups.

The order of difficulty of letters was determined by Keller from data secured on the learning of radio signals. He has recently revised this order on the basis of further data, but it was the best available at the time. How closely the relative difficulty of radio signals corresponds with that of blinker signals is not known, so far as I am aware.

Subjects came to the experimental room in groups of two to eight. The blinker, mounted on the wall, consisted of a 10-watt Mazda lamp in a Brownie Safe-light housing with a ground-glass window in the round opening. It was connected with a telegraph key which the experimenter operated by inserting his hand in an insulated box; this arrangement effectually deadened the click of the key. The experimenter attempted to keep the signals themselves at a constant rate of about six words per minute and to change his sending rate, when desired, by varying the intervals. These he controlled by subvocal counting, checked by a stopwatch. It is probable, however, that as the intervals were reduced the signals were sent at a slightly faster rate.

Each method was introduced by preliminary practice with number signals. Further details will appear in connection with specific experiments.

Preliminary Experiments

Experiment I: Dot-dash vs. Prompting Method

Specific Procedure

Subjects: Twelve *S*s were divided into two groups of six each.

Material: Twenty letters were divided into two equated lists of 10 each. The four easiest letters, *E*, *T*, *S*, and *N*, were omitted, as were *R* and *V*, owing to the widespread familiarity of the latter. The lists were as follows:

List I: A B C D G K L M X Y
List II: F H I J O P Q U W Z

Procedure: Each *S* came for two sessions 48 hours apart: in the first, he learned List I by one of the methods; in the second, List II by the other, according to the diagram in the previous section.

Method D consisted of five cycles of 200 sec. each, including rest pauses; 90 sec. were spent in study of the printed list and 60 sec. in presenting and correcting the test. Method P consisted of 10 trials, each including 10 signals and taking 70 sec., with a 30 sec. interval between trials, preceded by an initial presentation of the same length. The total learning periods were thus: for Method D, 16 min. 40 sec.; for Method P, 18 min. 20 sec. Each learning period was followed by a test consisting of 50 unprompted signals in random order.

*S*s were provided with record blanks marked off in double rows of squares, each row providing space for one trial.

Instructions: The instructions for the two methods, read and heard by each subject, were as follows:

Method D

"I shall give you a list of the 10 characters you are to learn with their corresponding signals in dots and dashes. A dot is represented by a short flash and a dash by a long one. When I say 'Start,' turn the list up and study it for 90 sec.; when I say 'Stop,' turn it down and be ready for the first test. The signals will be flashed in random order. Write the characters in the upper row marked '1' on your record blank. If you miss one leave the space blank. Next we shall correct the test. As I call the character, correct any errors or omissions by putting the correct character in the appropriate square of the lower row. Fold under the used part of the blank. You will then be started on another study period and repeat the same procedure until told to stop."

Method P

"The characters are represented by long and short flashes, each in a different grouping. First I shall name the 10 characters you are to learn and flash the signal that stands for each. Try to learn these pairs. Next I shall present a signal and after a few seconds name its character. Try to put the character in its box in the upper row marked '1' on your record blank *before* I name it. If you cannot, or if you make an error, put the correct character just below in the lower row *after* I name it. Do the same with the other signals as they are presented in random order. Keep trying to learn the pairs so as to fill the upper row correctly. After each trial, or series of 10 signals, count your errors, fold under the used part of the blank, and be ready for the next trial. Continue in this way until told to stop."

Rate of Sending: All final tests, and the practice tests in Method D, were presented at the rate of 17 letters per min., *i.e.*, with an interval of about 2½ sec. between signals. In Method P announcement of a letter occurred in the middle of an interval of approximately 6 sec.

Results

The final test, with a maximum score of 50, yielded the following means:

By Method D	27.6
By Method P	21.9
Difference in favor of Method D	5.7
t-ratio	1.7
Level of significance	10 percent

Eight of the 12 subjects scored higher by Method D, and 3 by Method P, while 1 scored the same by both.

Although the difference obtained was not reliable, it suggested a superiority of the conventional, dot-dash learning over the newer prompting method. The question arose whether a combination of the two methods might not preserve certain advantages of both. This possibility led to the next experiment.

Experiment II: Dot-dash vs. Combined Method

Specific Procedure

Subjects: There were 18 *S*s, divided into two groups of nine each.

Material: Same as in Experiment I.

Procedure: As in Experiment I, the experiment was conducted in two sessions 48 hours apart, and employed a system of rotation. Method D was duplicated in every respect. Method C consisted similarly of five cycles of 200 sec. each, including 90 sec. of studying the list and 70 sec. of prompting. The total learning periods were identical, 16 min. 40 sec., for the two methods. Both were followed by a 50-signal test.

Instructions: The instruction for Method C were as follows:

"I shall give you a list of the 10 characters you are to learn with their signals in dots and dashes. A dot is represented by a short flash and a dash by a long one. When I say 'Start,' turn the list up and look at the first character and signal. I shall name the character and flash its signal, then the next, and so on through the list. Try to learn these pairs. Continue to study the list until I say 'Stop'; then turn it down and be ready for Trial 1. In the trial, I shall present a signal and after a few seconds name its character. Try to put the character in its box in the upper row marked '1' on your record blank *before* I name it. If you cannot, or if you make an error, put the correct character just below in the lower row *after* I name it. Do the same with the other signals as they are presented in random order. Try to correct each error mentally as well as on the blank. After each trial count your errors and fold under the used part of the blank. You will then be started on another study period and repeat the same procedure until told to stop."

Rate of Sending: Same as in Experiment I.

Results

The mean scores in the final test were as follows:

By Method C	32.4
By Method D	28.5
Difference in favor of Method C	3.9
t-ratio	1.5
Level of significance	15 percent

These averages are based on 16 cases, since one *S* failed to finish one test. This necessitated discarding the scores of the most closely matched case from the other group also. Of the 16 *S*s, 9 scored higher by Method C, 5 by Method D, and 2 made equal scores. The above difference, although in the expected direction, was too small and inconsistent to be considered significant.

TABLE I

MEAN SCORES IN SUCCESSIVE TRIALS OR SUBTESTS IN EXPERIMENTS I AND II

Exp.	Method	Subtests or trials									
		1		2		3		4		5	
		1	2	3	4	5	6	7	8	9	10
I	P	2.0	2.8	4.3	4.7	4.5	4.9	5.4	5.1	6.1	5.8
	D		3.5		3.7		5.1		6.4		6.6
II	D		3.1		3.3		4.4		5.5		6.0
	C		3.0		4.8		5.3		6.6		7.3
	Mean		3.1		4.1		4.9		5.9		6.4

In order to determine the course of improvement in these experiments we have the scores in the 10 trials by Method P and in the 5 trials or subtests by Methods D and C. The scores, representing the number of correct responses in 10, were averaged and appear in Table I. Since the trends of all four series were closely similar, they were combined by averaging the even-numbered trials by Method P with the corresponding measures by the other methods. These means yield the negatively accelerated curve shown in Fig. 1.

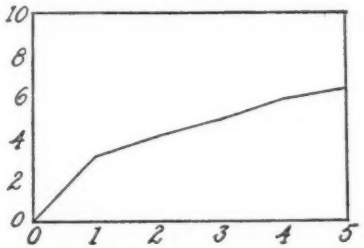


FIG. 1. Mean scores in successive trials or subtests in Experiments I and II. Data from Table I.

Experiment III: Method P vs. Method P′

The prompting technique was obviously not showing the superiority anticipated. It occurred to me as one possible reason that by the time a letter was announced the *S* had already forgotten the signal, and was therefore unable to profit by the information. Both because of a possible bearing on the controversy over the law of effect (**9**, **11**), and as a practical measure, it seemed worth while to try out a variation of the prompting method. The variation consisted simply of flashing each signal a second time immediately after announcing the letter it stood for. The following experiment was accordingly run to compare the original prompting method (P) with the revised method (P′).

Specific Procedure

Subjects: The 18 *S*s were divided into two groups of nine each.

Material: The same two lists of 10 letters each were used as in Experiment I, except that *R* was substituted for *I* in List II.

Procedure: The same system of rotation as in the previous experiments was used, covering two sessions 48 hours apart. Two modifications were introduced, however, in the technique of sending and receiving. The first consisted of increasing the length of a trial from 10 to 25 letters, presented in five groups of five each. This change was made to conform to conventional methods of instruction. The second and more important change was devised to eliminate the 'head-bobbing' which sometimes resulted in missing a signal. Instead of requiring the *S* to place each response in a separate box, he was provided with a new record sheet containing a number of large rectangles intended to serve merely as rough guides. He was told to use one rectangle for each group of five signals, and to look down only between groups.

By Method P, five trials were given in each session, lasting approximately 170 sec. apiece and separated by 20 sec. rest pauses. By Method P,′ owing to the repetition of each signal, a trial lasted about 210 sec., so that only four trials could be given in the same total time. The test which concluded each session, given after approximately 16 min. of learning, was lengthened from 50 to 100 signals presented in groups of five.

Instructions: The instructions for Method P were modified from Experiment I only with reference to the changes described under *Procedure*. The instructions for Method P′ were the same as for Method P with the insertion of the following sentence: "Keep watching the blinker constantly, because immediately after naming the character I shall repeat the signal to give added opportunity for learning."

Rate of Sending: In Method P, as in the previous experiments, signals were separated by approximately 6 sec. with prompting in the middle. In Method P′, prompting consisted both of naming the letter and repeating the signal; this occurred in the middle of a 7½ sec. interval, leaving the pauses before and after unchanged.

Results

The mean numbers of correct responses in 100 on the tests were as follows:

By Method P	53.6
By Method P′	52.1
Difference in favor of Method P	1.5
t-ratio	0.4
Level of significance	70 percent

Nine of the 18 *S*s scored higher by Method P, 7 by Method P′, and 2 scored the same by both. It was clear that repeating each signal after announcing the letter did nothing to enhance the effectiveness of the prompting method.

Two reasons may be advanced for this failure. First, the repetition of each signal cut down the number of trials which could be given in the allotted time. Secondly, to judge from *S*s' reports, the repetition was often confusing. The *S* was not always sure whether he was seeing the old signal or a new one. What was gained in impressiveness was probably lost by running the signals together. Method P′ was therefore dropped from further consideration.

Main Experiment

The foregoing experiments had failed to reveal dependable differences among the learning methods employed. Before accepting this negative conclusion as final, it seemed advisable to subject the three principal methods to a more thorough test. Besides using a larger number of *S*s, it was important to carry the learning beyond the initial stages and to compare the methods at higher rates of sending. By this means I hoped to test the belief which seems widespread among instructors in code, that students should be taught from the very beginning only those associations which they will use at the end. According to this view, any intermediate steps in translation, such as the use of dots and dashes, may become positively harmful when the student tries to put on speed. It was possible, then, that our

methods might show a difference only at a later stage of learning. This problem of the role of intermediate reactions in association is of some theoretical interest as well. A more inclusive experiment was therefore designed as follows.

Specific Procedure

Subjects: The 33 *S*s who completed the experiment were divided into three groups of 11 each.
Material: The 24 letters (all but *E* and *T*) were divided into three equated lists of eight each, as follows:

List I: B I K M P R U X
List II: A D G L Q S V W
List III: C F H J N O Y Z

Procedure: In order to compare Methods D, C, and P simultaneously, each group learned all three lists, each by a different method, as illustrated in Table II.

TABLE II

ORDER OF USE OF LEARNING METHODS

Group	List		
	I	II	III
A	Method D	Method C	Method P
B	Method C	Method P	Method D
C	Method P	Method D	Method C

With few exceptions, the *S*s came twice a week for three weeks. The two sessions of each week, 48 hours apart, were devoted to the learning of a single list. Five cycles were completed on each list, two in the first session and three in the second. A cycle, it will be recalled, consisted of a prescribed period of study or prompting or both, followed by a test. In successive cycles, the study or practice periods became progressively shorter and the rate of sending progressively faster, as noted below. Following is the complete time schedule for the three methods:

TIME SCHEDULE

Day I

Method D	Method C	Method P
0:00 Study	0:00 Presentation	0:00 Presentation
5:00 Testing	0:30 Study	0:40 Prompting
10:00 *Test*	5:00 Prompting	10:00 *Test*
16:00 Study	10:00 *Test*	16:00 Prompting
20:00 Testing	16:00 Study	25:00 *Test*
25:00 *Test*	20:00 Prompting	
	25:00 *Test*	

Day II

0:00 Study	0:00 Presentation	0:00 Presentation
3:00 Testing	0:30 Study	0:40 Prompting
7:00 *Test*	3:00 Prompting	7:00 *Test*
12:00 Study	7:00 *Test*	12:00 Prompting
14:00 Testing	12:00 Study	18:00 *Test*
18:00 *Test*	14:00 Prompting	23:00 Prompting
23:00 Study	18:00 *Test*	28:30 *Test*
25:00 Testing	23:00 Study	
28:30 *Test*	25:00 Prompting	
	28:30 *Test*	

The modifications of technique introduced in Experiment III above were retained. Practice trials consisted of 25 letters in groups of five. Recording was done on the sheets containing large rectangles, without looking down. The five tests, italicized in the time schedule, which provided the data for comparison, were reduced to 50 signals.

In making up the presentation schedules for trials and tests, one additional change was introduced in the interest of economy of learning. In each group of 25 signals the three most difficult letters appeared four times each, the three of intermediate difficulty three times each, and the two easiest letters twice each.

Instructions: The instructions for Method P were practically identical with those in Experiment III. The instructions for Methods D and C were similarly derived from those used in the first two experiments.

Rate of Sending: The tests on Day I were sent at the rate of three words per min. Tests in the first two cycles on Day II were sent at $4\frac{1}{4}$ w.p.m. and in the final cycle at $5\frac{1}{4}$ w.p.m. In prompting trials, the corresponding intervals occurred both before and after prompting, except for a lag in the final cycle to permit subjects to record and correct their responses.

Results

In order to make an over-all comparison of the three methods, the scores on all five tests by a given method were averaged. The means of the three groups appear in Table III. When the three means for

TABLE III

MEAN NUMBER OF CORRECT RESPONSES, ALL TESTS

Group	Method		
	D	C	P
A	28.0	33.7	37.7
B	30.3	25.1	29.7
C	29.2	30.6	22.0
Mean	29.1	29.8	29.8

each method were averaged, they gave the final means appearing at the foot of the table. The differences are practically negligible. Two of the methods, C and P, have identical means at 29.8, while Method D is only 0.7 below them. To make certain, however, that the means did not vary significantly, the null hypothesis was tested

TABLE IV

TESTING THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN METHODS BY ANALYSIS OF VARIANCE: DATA FOR COMPUTATION OF F

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sq. (Variance)
Methods	2	1.02	0.51
Ss' means	32	4516.73	141.15
Interaction	64	7548.91	117.95
Total	98	12066.66	

$F = 0.51/117.95 = .004$

by analysis of variance (1). The essential computations are presented in Table IV. Dividing the variance between methods by that of the interaction gives an F of .004, which clearly fails to reject the null hypothesis.[1] As measured, the three methods may be considered equivalent.

Nevertheless, the possibility remained that the combined scores might obscure differences which a finer analysis would reveal. Table III, for example, shows a considerable practice effect within groups.

TABLE V

WEIGHTED MEANS ARRANGED IN ORDER OF LEARNING

Group	List I		List II		List III	
	M	σ_m	M	σ_m	M	σ_m
A	(D)* 28.0	±1.8	(C) 33.7	±1.6	(P) 37.7	±1.5
B	(C) 29.4	±2.0	(P) 34.7	±2.1	(D) 35.5	±2.0
C	(P) 26.9	±1.5	(D) 35.5	±1.6	(C) 37.3	±1.5

* Letters in parentheses indicate methods used.

It suggests the possibility that, although the three methods are equal in the long run, one method may be superior at an early stage of practice and another at a later stage. To make this comparison the groups should have been equated at the start. Since they were not, the only alternative was to compensate for the differences statistically. A comparison of the total scores for the entire experiment showed that the group means could be equated if that of Group A was multiplied by 1, Group B by 1.17, and Group C by 1.22. The group means for each method were accordingly multiplied by these weights in Groups

TABLE VI

MEAN NUMBER OF CORRECT RESPONSES IN SUCCESSIVE TESTS

Method	Tests 1 & 2	Tests 3 & 4	Test 5
D	28.2	31.4	26.4
C	28.6	32.1	27.6
P	28.4	32.0	28.2

A, B, and C, respectively. Arranged in order of learning, the weighted means appear in Table V.

A glance at the table shows that our supposition is not borne out. True, Method P ranks lowest on List I, in the middle on List II, and highest on List III. But the intra-list differences, as compared with the standard errors, are all too small to be dependable. Within

[1] To be significant at the 5 percent level, F must equal 3.14 (6).

the limits of this experiment general practice, or list-to-list transfer, is not an important factor in differentiating the methods of learning.

A second variable which might be related to learning method was rate of sending. One method might excel at slower rates and another at faster. As described above, Tests 1 and 2 were sent at 3 w.p.m., Tests 3 and 4 at $4\frac{1}{4}$ w.p.m., and Test 5 at $5\frac{1}{2}$ w.p.m. Table VI shows the average scores by the three methods for each group of tests. The differences between these scores appear in Table VII. A pro-

TABLE VII

MEAN DIFFERENCES BETWEEN METHODS IN SUCCESSIVE TESTS

Methods	Tests 1 & 2	Tests 3 & 4	Test 5	
			M σ_m	t
P–D	0.2	0.6	1.8±1.65	1.09
C–D	0.4	0.7	1.2±1.37	0.88
P–C	−0.2	−0.1	0.6±1.70	0.35

gressive increase is to be noted in the superiority of both the prompting and combined methods over the dot-dash method. If significant, this trend might be attributed either to the increase in rate or to specific practice with a given list of signals. As shown in the last column of Table VII, however, even the largest difference is not significantly greater than its standard error.

A third possibility is that one method might be better adapted to fast learners and another to slow learners. To answer this question, the three *S*s with the highest scores were selected from each group, and the scores of these nine were averaged for the three methods separately. The same was done with the three lowest scorers from each group. The results appear in Table VIII.

TABLE VIII

MEAN SCORES OF THREE HIGHEST AND THREE LOWEST SCORERS FROM EACH GROUP

Group	Method D	Method C	Method P
High	42.0	41.5	40.3
Low	18.0	18.7	19.4

The differences between methods for both high and low groups are small. Analysis of variance yielded *F*-ratios of 0.37 and 0.17, respectively, giving no evidence of significant deviations. It is noteworthy, however, that, taking the methods in order from D through C to P, the fast learners show a progressive decrease of score and the slow learners a progressive increase. The regularity

of the trends suggests that the good learners were able to make slightly more efficient use of dot-dash equivalents in code reception than the poor learners.

A final question concerns the learners' own reactions to the three methods. At the end of the experiment, the *Ss* were asked to rank the methods in order of preference. The frequencies with which each method was put in first, second, or third place appear in Table IX.

TABLE IX

FREQUENCIES WITH WHICH METHODS WERE RANKED 1ST, 2ND, OR 3RD IN ORDER OF PREFERENCE

Group	A (9)*			B (11)			C (11)		
Rank	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
Method									
D	0	4	5	1	5	5	2	8	1
C	2	4	3	1	5	5	8	3	0
P	7	1	1	9	1	1	1	0	10

* Figures in parentheses indicate number of *Ss*. Two *Ss* in Group A failed to make a rating.

The ratings vary widely from group to group. In Groups A and C there is an inverse relation between the methods' ranks and the order in which they were used. Method P, for example, was the last one used by Group A and the first by Group C. In the former group it was rated first by 7 out of 9 *Ss*; in the latter it was rated third by 10 out of 11. The obvious inference is that the *Ss* tended to prefer the method preceded by the most practice; in other words, they were estimating their own performances rather than the methods used. Group B, however, fell out of line with this interpretation; Method P, which came second, was preferred by 9 out of 11, while Method D, which came last, was liked no better than the one used first. It is worth noting that Group B showed less improvement from list to list than the other two.

DISCUSSION

The present experiment was undertaken with the theoretical expectation that the prompting method would prove definitely superior to the dot-dash method of code-learning. The expectation was based on three principles: (1) that the stimulus patterns responded to in practice should be the same as those encountered in the task itself; (2) that recitation is more effective than reading; (3) that knowledge of results should be provided in the form of an immediate check-up after each response. Experimental evidence in support of each of these principles could readily be adduced (4).

It would seem that the prompting method holds the advantage on all three counts. Aside from the difference between marks on paper and a flashing light, Morse code alphabets are usually printed with the signal *after* the letter, whereas in receiving code it comes first. The student *may* reverse this order for himself, just as he *may* resort to self-prompting; but without special instruction or incentive he may not. Again, in the training methods commonly used, the student has no opportunity to check his accuracy until at least 25 signals have come and gone.

The negative results of the experiments came as a surprise. They obviously call for a reëxamination of the two methods.[2] A closer analysis, based partly on observations and *S*s' comments in the course of the experiments, reveals a number of advantages for Method D previously overlooked:

1. Dot-dash learning affords closer contiguity between the symbols to be associated. Letter and signal can be grasped as a single unit. In the prompting method, enough time must elapse between signal and letter to allow for writing a response. *S*s complained that by the time the letter was announced, they had forgotten the signal. This objection might be overcome, as Dr. Keller suggested to the writer, by repeating the signal *during* the interval. But it is doubtful if even this addition would provide as close an integration as the other method.

2. Dot-dash learning facilitates the noting of relations among the signals. Much of the initial difficulty of code learning is due to confusion among similar patterns. It seems likely that the confusion can be lessened by bringing patterns together and noting both their similarities and differences. This argument is weakened by Taylor's recent evidence (**8**) that initially presenting similar characters together produces no faster learning than dissimilar characters. Nevertheless, the experimental investigation of memorizing has shown that the process is typically one of constructing an integrated pattern rather than forming separate associations (**12**). The dot-dash method leaves the student free to manipulate his materials at will, bringing various associates into juxtaposition. The prompting method is far more rigid in this respect. If the student has difficulty in remembering the signal just presented, he has relatively little chance to compare that signal with any other.

3. The same relative freedom of Method D enables the learner to use his time more economically. He can devote more time to the hard letters and less to the ones already learned. An attempt was made in the main experiment to achieve the same end by the prompt-

[2] Method C, which was simply a combination of the other two, is here left out of consideration.

ing method. Hard letters were presented twice as frequently as easy ones. But this device, based on the *average* difficulty of letters, is a shot-gun method with no individual precision.

4. In Method D the practice tests interspersed among study periods were given under identical conditions with the final tests of achievement. In Methods P and C, on the other hand, prompting injected a foreign element and increased the time between signals.

With the exception of the last point, the advantages ascribed to Method D are such as the learner may or may not make use of. Whether or not he does so will presumably depend on his motivation, alertness, intelligence, experience, and instructions. Here is a possible basis for the tendency noted for fast learners to do relatively better by Method D than slow learners. But in general our analysis indicates that both methods can find support in learning theory. It suggests that the lack of clear-cut difference in our results reflects an actual balance between the assets and liabilities of the methods.

The present findings point to the following recommendations for the initial mastering of the International Morse Code by flashing light. Students may be given printed dot-dash alphabets for preliminary study without fear of harmful after-effects. They should be given special instructions in self-testing, making careful discriminations between confusing signals, and concentrating on difficult items. Their study should be supplemented early by practice in receiving blinker signals using the prompting method. It is not to be expected that transfer from lists to blinker will be perfect, but it should be substantial, and certainly positive rather than negative. However, the value of list study falls off as the student progresses, so that time should be increasingly given to direct practice in reception. As soon as errors have been reduced to a fairly low level, the prompting method may be supplemented in turn by mixed copy without prompting.

It is not within the writer's province to make more specific recommendations as to teaching procedures. Certainly the data of the present experiments do not warrant them. It is more appropriate to raise two questions in an attempt to evaluate these data:

1. How safely can our results be generalized to sound reception, *i.e.*, to radio telegraphy? Although the question cannot be answered without experiment, certain differences may be noted which stand in the way of generalizing. In the first place, auditory signals are presented much more quickly than visual, giving even less opportunity to 'count dits and dahs.' In the second place, radio operators attain much higher rates of sending and receiving than is possible by blinker. It seems extremely doubtful, however, that the methods used at the beginning of learning would continue to exert a differen-

tial effect at such advanced stages. In the third place, auditory signals may be easier to 'hold in mind' until the characters are announced, and may thus be more adaptable to the prompting method than visual stimuli. These characteristics of radio reception might be expected to favor the prompting method. Concerning a recent series of experiments in teaching radio telegraphy (**8**), however, Dr. D. W. Taylor wrote me as follows: "A method of immediate reinforcement . . . produced no faster learning than the usual method of delayed reinforcement."

2. How safely can our results be generalized to learning the entire 36 characters? In these experiments only 8 or 10 letters were learned by any one method. This restriction was necessitated by the rotation system used. A more adequate experiment would compare equated groups in learning all the characters, each group using a different method. This plan would have the further advantage of permitting the methods to be compared stage-by-stage in the course of learning, instead of combining earlier and later stages in a single average based on widely varying scores. Unless large numbers of Ss were available, however, the groups would have to be matched on the basis of tests for code-learning aptitude. The development of such tests, which is proceeding rapidly at present, should thus pave the way to the experimental discovery of optimal training procedures.

In stating the problem of the main experiment, I raised the theoretical question of the value of intermediate reactions in memorizing. It is clear that the present data are confined to too restricted a range to justify a general conclusion. At the same time it is worth noting that a *general* conclusion is not to be expected, at least in the near future. The modern emphasis is on the relativity of performance to many variable factors (**4**). Only when a large number of conditions—*e.g.*, type of material, kind of intermediate process, method of measurement, ability of subject—have been varied independently and together can we state the essential relationships.

Summary

A series of experiments was performed to compare methods of conducting the initial stages of learning to receive International Morse Code by flashing light. The following results were obtained:

1. A prompting method of presenting blinker signals was not found superior to a method of studying printed lists of dot-dash equivalents.

2. A method combining dot-dash study with prompting was not found superior to either of the two used separately.

3. No dependable differences were found among the three methods at different stages of practice or rates of sending.

4. Repeating a blinker signal immediately after prompting was without effect on rate of learning.

5. There was a slight but undependable tendency for the slower learners to do relatively better with the prompting method and for the faster learners to do relatively better with the dot-dash method.

(Manuscript received April 28, 1943)

References

1. Garrett, H. E., & Zubin, J. The analysis of variance in psychological research. *Psychol. Bull.*, 1943, **40**, 233–267.
2. Keller, F. S. An improved procedure for teaching International Morse Code reception. Unpublished mimeographed material, 1942.
3. Koch, L. Arbeitspsychologische Untersuchung der Tätigkeit bei der Aufnahme von Morsezeichen, zugleich ein neues Anlernverfahren für Funker. *Z. angew. Psychol.*, 1936, **50**, 1–70.
4. McGeoch, J. A. *The psychology of human learning.* N. Y.: Longmans, 1942. Pp. 633.
5. Peak, H. Koch's method of learning code reception. *Psychol. Bull.*, 1942, **39**, 495.
6. Snedecor, G. W. *Statistical methods.* (2nd ed.) Ames, Ia.: Collegiate Press, 1938. Pp. 388.
7. Taylor, D. W. Learning telegraphic code. *Psychol. Bull.*, 1943, **40**, 461–487.
8. Taylor, D. W. The learning of radiotelegraphic code. *Amer. J. Psychol.*, 1943, **56**, 319–353.
9. Thorndike, E. L. The direct action of rewards upon mental connections and their indirect action via the stimulation of inner equivalents of the connections. *J. exp. Psychol.*, 1935, **18**, 91–96.
10. Trabue, M. R. Psychology and the war: The subcommittee on learning and training. *Psychol. Bull.*, 1943, **40**, 57–59.
11. Wallach, H., & Henle, M. An experimental analysis of the law of effect. *J. exp. Psychol.*, 1941, **28**, 340–349.
12. Woodworth, R. S. *Experimental psychology.* N. Y.: Holt, 1938. Pp. 889.

THE OCCURRENCE OF PLATEAUS IN TELEGRAPHY

BY HOMER B. REED AND HARVEY A. ZINSZER

Fort Hays Kansas State College

The concept of plateaus in learning was brought into prominence in psychology by the much-quoted experiment of Bryan and Harter (1) on learning telegraphy. Although Bryan and Harter's experiment provoked an immense amount of investigation and publication in the field of sensory-motor learning, the interesting field of telegraphy was ignored. The writers have been unable to find any published experiments on telegraphy since Bryan and Harter's in 1897 and 1899.[1] Trabue (3), however, reports that investigations have been carried on by the Signal Corps, by numerous civilians, and in particular by Dr. Fred S. Keller of Columbia University, but these have not yet been published. Although investigators of sensory-motor learning other than telegraphy have not found the plateau to be universal in learning, (*cf.* 2) there has hitherto been no evidence against its existence in learning the reception of the telegraphic code.

In Bryan and Harter's report there appear the receiving curves of ten learners, all of whom show an extended plateau occurring before the minimum main line speed of 72 letters per minute is reached and continuing from ten to twenty or more weeks. If this plateau could be eliminated, the training period for this important occupation could be greatly reduced in length. The main problem of this study was to reinvestigate the general form of the learning curve for the reception of the telegraphic code, giving special attention to the possibility of eliminating the plateau. Our theory was that this might be done by spending most of the teaching time on receiving, and by teaching receiving and sending at different times, rather than simultaneously. Accordingly, our plan was to distribute the teaching time as follows: one month on receiving, one week on sending, a second month on receiving; and, for the remaining seven weeks of the semester, three fourths of the time on receiving and one fourth on sending. We believed that the appreciation gained through receiving, of the proper time intervals for dots, dashes, and spaces between words and letters would make possible the quick learning of sending. This belief proved to be correct.

[1] Since the above was written an excellent review of the literature, published and unpublished, has been written by Donald W. Taylor in *Psychol. Bull.*, 1943, 40, 461–488.

Procedure

An excellent opportunity to make a further study of telegraphic learning occurred when our college organized two classes of 24 students each in Practical Radio to satisfy an urgent demand of the Government for trained men in this field. This was done during the spring semester of 1942.

The layout of the equipment consisted of six tables 40 in. wide, 6 ft. long, and 30½ in. high, arranged in the form of a U. A 1000 cycle General Radio type 213-Audio-Oscillator energized by a 6-volt Edison storage battery provided the necessary signal hum. The hum from this oscillator was transmitted along both branches of the U system of tables by means of a lamp cord, which served as a transmission line. On each table, and connected in parallel across the transmission line, were two wooden instrument boards about 12 in. by 18 in. by ½ in. The following equipment connected in series was mounted upon each of these boards: two binding posts for connecting to the transmission line; a telephone condenser of either one or one-half microfarad capacity; a telegraph transmitting key; and a pair of dual or triple Fahnstock clips for connecting the telephone headsets into the circuit. The condensers prevented the line from being shunted whenever one of the keys is depressed. In this manner one to three students could work at any one of the instrument boards.

Whenever it was desirable to transmit to the class as a whole, a transmitting key, a Gardiner-Levering Type S Automatic Sender, inserted in series with the main transmission line, was available for this purpose. The student sections met four 50-minute periods per week for supervised practice in receiving radio signals.

The men were tested once a week for a two-minute period, and the results were scored in terms of letters per minute. Five letters were counted as one word. For each test new sentences of approximately equal difficulty were used. At the end of the semester the records of 43 subjects were available for use.

Results and Discussion

The character of the results may be seen from Table I and Figs. 1 and 2. Fig. 1, based on the average of Table 1, shows an S-shaped curve. Below it, for the sake of comparison, is shown the curve of Bryan and Harter's subject, B.L., for the first 16 weeks. Some of the differences that appear are the following:

1. Our curve shows positive acceleration and slow initial progress, whereas Bryan and Harter's curves show negative acceleration and rapid initial progress.

TABLE I

Means, Standard Deviations, Maxima, and Minima Obtained in the Semester Course in Receiving Telegraphic Code. (Score equals letters per minute)

Week	1	2	3	4	5	6	7	8	9*	10	11	12	13	14	15	16
Number Tested	27	30	34	41	40	40	40	40		43	22	35	20	40	41	42
Maximum Score	25	30	35	50	55	70	75	75		80	80	80	80	80	85	85
Mean	18.1	18.0	22.2	26.1	36.1	41.5	48.6	50.6		54.2	61.6	61.6	61.0	64.2	65.9	64.4
S.D.	5.6	5.7	6.5	8.0	9.8	11.3	10.8	11.1		11.9	10.5	8.6	11.0	12.5	11.5	12.7
Minimum Score	10	10	15	15	20	20	30	30		35	40	40	40	35	45	45

* No tests were given at the end of the ninth week because of mid-semester examinations in other subjects.

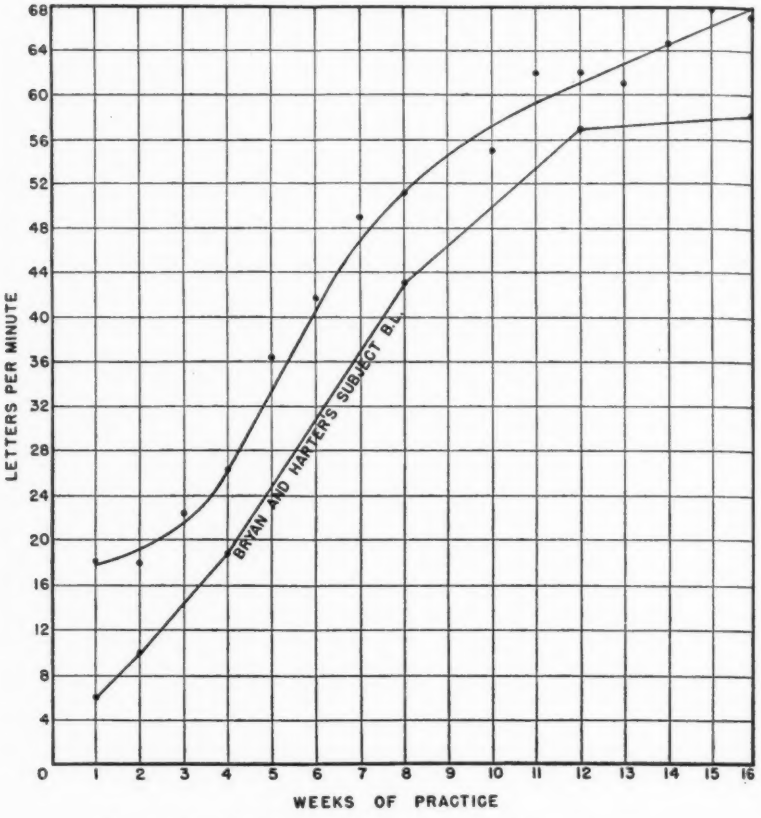


FIG. 1. Learning the telegraphic code. Averages of 20–43 learners.

2. Our curve shows no marked plateau, whereas all of Bryan and Harter's subjects, ten in number, show marked and extended plateaus for receiving.

3. Our curve shows more rapid progress. The curves of the subjects given by Bryan and Harter all fall below our average.

At first thought one might suppose that the S curve is the typical learning curve; but this view is given up after we look into the conditions of the experiment and at the individual learning curves. Conditions in this experiment which favor the initial positive acceleration are the novelty of the learning, the complexity of the material, the slight possibility of transfer from previous learning, and the high ability of the learners. In conformity with these conditions, the majority of our subjects show positive acceleration, a fact that is

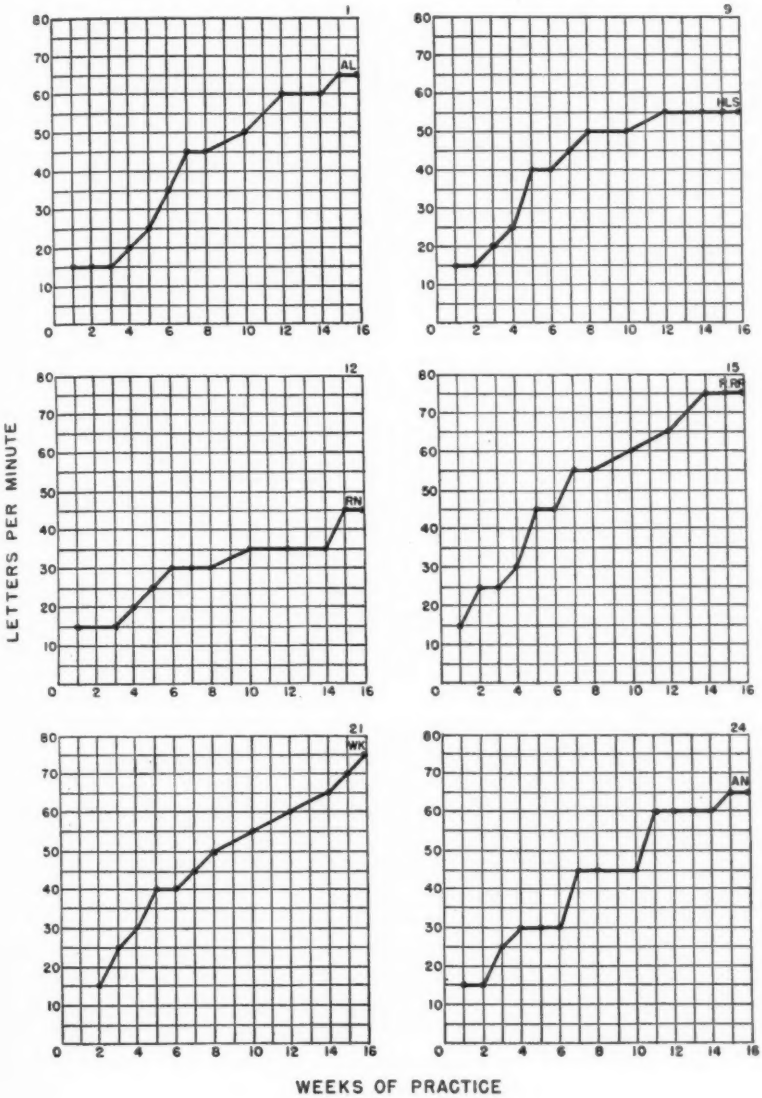


FIG. 2. Individual learning curves for receiving the telegraphic code.

illustrated in the individual curves for Subjects 1 and 9 in Fig. 2. The trait, however, is not universal; and this fact is illustrated in the individual curves for Subjects 12, 14, 21, and 24. The fact that none of Bryan and Harter's subjects show initial positive acceleration is strange, but this may be a matter of selection, or it may be due to the conditions and methods of instruction.

Extended plateaus among our learners are the exception rather than the rule. This may be thought to be due to the fact that our learning period was too short to permit them to appear, for in Bryan and Harter's subjects they appeared between the 12th and 17th weeks, and continued for periods of ten to twenty weeks. In all cases they appeared before the subject reached the minimum main line rate of 72 letters per minute. But this explanation is unlikely, for 12 of our subjects reached or exceeded 75 letters per minute without a prolonged plateau, and two of them without any plateau. One of these is illustrated in the curve for Subject 21 in Fig. 2. If an extended plateau means little or no progress for ten weeks or longer, then we had only one subject who approached this. Subject 10 made rather rapid progress up to the end of the seventh week, when he reached a speed of 45 letters per minute, a speed which he kept for the remaining nine weeks of the semester. None of our other subjects shows any such long period without progress. But many of them do have short periods of no progress, such as may be seen in most of the curves in Fig. 2. Whether or not they should be called plateaus is a matter of definition. If we define a plateau as no measured progress for two weeks or three successive tests, 13 of our subjects have no plateaus. If we define it as no measured progress for three weeks or four successive tests, 29 subjects show no plateaus. But if we define a plateau as no measured progress for two successive tests one week apart, all of our subjects have plateaus, although this is not a plateau as the term was used by Bryan and Harter. In view of these facts, the plateau cannot be said to be universal.

If we define a plateau as no measured progress for two weeks of regular practice, it is a point of interest to notice when it occurs among our subjects. Three such periods occur at the beginning of practice, as in the curve for Subject 1, Fig. 2. Twelve occur during the middle range of practice, as in the curve for Subjects 12 and 24, and 18 occur at the end of practice, as in the curve for Subject 9. This last fact is consistent with the time of appearance of extended plateaus in Bryan and Harter's subjects. The fact that plateaus usually do not occur until considerable progress has been made is in agreement with the theory of Bryan and Harter that they are due to the presence of imperfect elementary habits, although other fac-

XUM

tors may be at work. The fact that they sometimes occur at the beginning may be due to the complexity of the initial habits for certain learners.

The more rapid progress of our subjects as compared with Bryan and Harter's may be due to one or more of the following factors: selection, motivation, and differences in method of presentation. Although the curves of Bryan and Harter's subjects all fall below our average, some of our subjects progressed at about the same rate as did theirs. The motivation added by the war demand, and the prospect of a good job at the end of a semester provided the required minmum rate was met, should have made some difference. It is also probable that more rapid progress is made when the teaching emphasis is principally upon receiving than when it is equally distributed between sending and receiving, as appears to have been the case with Bryan and Harter.

Summary

The problem of this experiment was to investigate the characteristics of the learning curve for telegraphy as found by Bryan and Harter, giving particular attention to the possibility of eliminating the plateau. Forty-three college students were given a semester's course in telegraphy. Three differences were found: (1) initial positive acceleration instead of negative acceleration, (2) no prolonged plateaus and in some cases none at all, (3) more rapid learning. These differences are attributed to the use of a larger number of subjects, and to differences in presentation, the important features of which were the separate teaching or receiving and sending, and devoting most of the time to receiving. The experiment was not completely comparable to the Bryan and Harter experiment, since it lasted only for sixteen weeks, whereas theirs continued for forty weeks.

(Manuscript received May 25, 1943)

References

1. Bryan, W. L., & Harter, N. Studies in the physiology and psychology of the telegraphic language. *Psychol. Rev.*, 1897, **4**, 27–53; 1899, **6**, 345–375. Reprinted in: Bryan, W. L., Lindley, E. H., & Harter, N. *On the psychology of learning a life occupation.* Bloomington, Ind.: Ind. Univ. Book Store, 1941.
2. Kao, D. L. Plateaus and the curve of learning in motor skill. *Psychol. Monogr.*, 1937, **49**, No. 219.
3. Trabue, M. R. The subcommittee on learning and training. *Psychol. Bull.*, 1943, **40**, 57–59.

GENERALIZATION IN THE INITIAL STAGES OF LEARNING NONSENSE SYLLABLES: II. PARTIAL AND INADEQUATE RESPONSES

BY B. R. PHILIP

Fordham University

AND

HELEN E. PEIXOTTO

Hunter College

Introduction

In a previous paper (7) we dealt with the initial stages of learning nonsense syllables, restricting ourselves to the integral or correctly reproduced responses whether or not they were properly localized. When defining our terms, we called the proper localization of integral responses, discrimination, and the improper localization of integral responses, generalization.

As a rule, in experimental work on verbal material, learning is measured exclusively by the correctly reproduced and localized response, *i.e.*, by the discriminated response. It is obvious that generalized integral responses also indicate some degree of learning, when, according to the definition of this study, the subject reproduces the syllable by spelling it correctly, although the response is not assigned to the appropriate stimulus syllable. Recent studies seem to indicate that generalization plays an important role in learning, forgetting, inhibition, reminiscence and such phenomena. But if the generalized response must be taken into account in the study of learning it seems important to pursue the matter further and to investigate not only the integral responses, whether discriminated or generalized, but also the partial and inadequate responses as defined in the previous paper. By partial responses are meant inverted responses and responses with one letter incorrect, whether properly or improperly localized. These responses give us partial discrimination and partial generalization respectively. While not subject to objective determination, a still lower degree of learning exists among the inadequate responses, which consists in the giving either of no response, or of a syllable with two or more letters wrong.

It is of course admitted that we are departing from the general usage when we take partial and inadequate responses to indicate some form

XUM

of learning. But it must be remembered that in this paper we are studying the initial phases of learning, during which the subject will necessarily make a large number of responses other than integral responses. Since some form of learning is undoubtedly present almost from the very beginning, the study of these partial and inadequate responses will at least give some information in regard to these initial stages of the learning process.

Procedure

The procedure used to obtain partial discrimination and partial generalization has been described fully elsewhere; that is, the procedures of experimentation and the methods of obtaining results are exactly the same as those detailed for integral responses in the previous paper (7). Generalization gradients according to position, and an average measure of partial discrimination and partial generalization, were obtained using, however, partial instead of integral responses as the basis for computation.

The method for computing inadequate responses is, of necessity, radically different from those described above for integral and partial responses. Inadequate responses, since they cannot be localized, cannot, of course, be analyzed for generalization to any particular stimulus. However, if the sum of the integral and partial responses at any particular position with reference to the basic syllable is deducted from the total number of possible responses, the number of inadequate responses corresponding to that particular position may be inferred. This procedure has been carried out for all the series.

Results

Generalization According to Serial Position.—The generalization curves for position for partial responses may be seen in Fig. 1. There is an initial gradient only for the 8-series up to the fourth

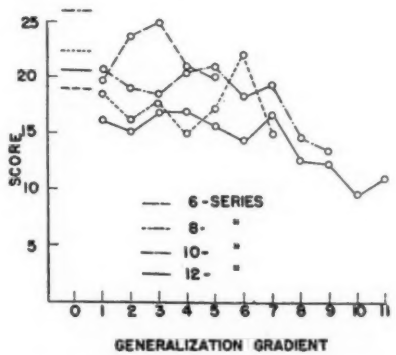


Fig. 1. Generalization gradient according to serial position of the syllables in the learning order, for partial values.

The ordinates indicate the total score. The abscissae give the generalization gradient according to distance from the basic (discriminated) syllable. Above position O, the lines, coded according to series, indicate discriminated values, the ordinate scale for which is one-third that of the generalized values. There is no clearly seen gradient, except for the final positions of the longer series, which are probably not reliable.

position. The gradients in the other three series seem to be deferred for at least three positions. At any rate, negative acceleration occurs for all series somewhere along the course of the curves. It may be that localization, particularly for the first few positions, is largely a matter of chance and is distributed fairly evenly over these positions; however, these positions are chosen more frequently for localization of the response than those more distant from the basic syllable. The partial generalization gradient is not nearly as apparent as the integral generalization grandient previously found, and, since it occurs in the end positions of the series, is not nearly so reliable.

The results for inadequate responses according to serial position are presented in Table I. When allowance is made for varying dif-

TABLE I

GENERALIZATION GRADIENT FOR INADEQUATE RESPONSES

This table is derived by distributing the non-localized responses uniformly among all positions except the basic position, and by subtracting, from the values thus obtained the generalized responses at each position.

This gradient is presented only for comparison with the other gradients, as, properly speaking, inadequate responses cannot be localized.

Positions	1	2	3	4	5	6	7	8	9	10	11
6-Series	149.7	142.1	149.5	154.7	148.2						
8-Series	137.1	139.0	137.5	141.2	142.5	127.4	145.2				
10-Series	166.2	169.5	173.1	173.1	166.8	172.2	171.5	177.4	171.3		
12-Series	145.6	147.1	145.3	146.4	148.1	149.9	147.8	149.6	149.0	144.2	151.0

ficulty, and associational and positional effects which occur chiefly towards the end of the series, it is apparent that the inadequate responses show no gradient for position. This is to be expected, as inadequate responses indicate almost no degree of correct perception and hence cannot be assigned by the subject, even approximately, to the correct stimulus. In fact the method here used of calculating inadequate responses is based on the assumption that responses which are not localized are distributed uniformly among all positions except the basic position. Table I is valid, therefore, only inasmuch as it presents the relative frequency of inadequate responses as compared with other responses.

Relationship Between Partial Discrimination and Partial Generalization.—In determining discrimination and generalization values for each syllable based upon partial responses the same technique was made use of as for integral responses. In this case a somewhat different order of discrimination values among the syllables was found, but there was a good agreement ($\rho = 0.81$) between the results no matter which order of discrimination was used in the comparison with partial generalization.

TABLE II

RELATION BETWEEN PARTIAL DISCRIMINATION AND PARTIAL GENERALIZATION

The values are given for syllables listed in descending order of partial discrimination. The frequency of occurrence of the syllable pair is given in column 2; thus, QOS–MEF appeared in all four series. The values are averaged for frequency of occurrence.

Syllables	Occurrence	Partial Discrimination	Generalization of Response	Generalization of Stimulus
MIB–QUJ	3	7.7	9.2	17.6
QIM–JEX	2	7.4	9.8	10.8
QOS–MEF	4	7.4	8.4	10.6
ZOF–GEX	4	5.9	10.3	15.5
GEC–VAF	4	4.8	11.0	8.5
QOJ–CEF	1	4.2	10.3	8.0
VUF–ZEQ	3	3.8	11.1	10.5
YIX–ZOJ	4	3.6	10.7	9.3
KOJ–YUF	1	3.0	9.6	8.5
ZUV–CIJ	4	2.2	10.9	9.8
YAV–WUQ	4	2.1	12.6	7.1
YIF–VUQ	2	1.0	10.4	7.4

Table II and Fig. 2 show that there is a direct agreement between partial discrimination and partial generalization of stimulus, and that there is an inverse relationship between partial discrimination and

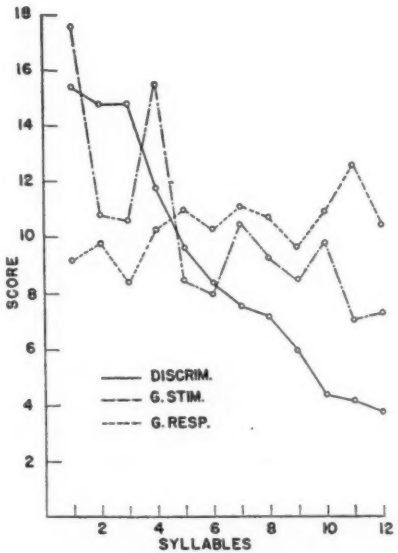


FIG. 2. Relation between partial generalization of syllables and partial discrimination.

The ordinates are the total scores, averaged for number of occurrences of the syllables in the four series. The abscissæ give the order of discrimination; the most frequently discriminated syllables are first in order. The scale values for discrimination are twice those for generalization.

partial generalization of response. This relationship between partial generalization of response and partial discrimination is barely significant, as will be seen in the next section, but it is interesting, since no relationship was obtained between the two variables when integral values were used. The direct agreement between generalization of stimulus and discrimination for partial responses is in accordance with the relationships found for integral responses and the same explanation for this agreement as previously given (7), in terms of relative inhibition of first and second members of the paired associates, seems equally applicable here.

It must be remembered that in this section we are dealing with the discrimination values of the syllables based upon averages for each individual syllable from all the series. The discrimination and generalization values differ somewhat from series to series, since the syllables were rotated in position, but the results in general are consistent.

Intercorrelation of Variables.—Based upon the total responses per syllable, and using the Spearman rank-difference method, an intercorrelation table was drawn up and is presented in Table III. As

TABLE III

INTERCORRELATIONS OF VARIOUS TYPES OF RESPONSE

Correlations calculated by the Spearman Rank Difference Method for each of the 12 syllables, the responses for which have been averaged for number of trials. Only values greater than 0.58 are statistically reliable. ID, Integral Discrimination; IG_r, Integral Generalization of Response; IG_s, Integral Generalization of Stimulus; PD, Partial Discrimination; PG_r, Partial Generalization of Response; PG_s, Partial Generalization of Stimulus.

	ID	IG_r	IG_s	PD	PG_r	PG_s
ID	———					
IG_r	−0.17	———				
IG_s	0.83	−0.35	———			
PD	0.81	−0.03	0.62	———		
PG_r	−0.58	0.31	−0.51	−0.59	———	
PG_s	0.40	−0.17	0.45	0.72	−0.45	———

the number of syllables varied with the series, (*viz.*, two syllables appeared only in the 12-Series; four syllables appeared only in the 10-Series and the 12-Series, etc.,) the responses were first averaged for frequency of occurrence, *i.e.*, whether they occurred in four, three, two or one series. As there were only 12 syllables in all, the reliabilities of the correlations were low, and only values of ρ greater than 0.58 are reliable at the 1 percent level of significance. Although only the results based upon the average of all four series are here presented, the same procedure was applied to each individual series, and the consistency of the rankings from series to series which was then found may be taken to indicate the trend of the results.

It is readily seen from Table III that there is a consistent agreement between the correlations of the integral and the partial responses. For both types of response, discrimination yields low, unreliable and negative correlations with integral generalization of response; reliable and negative correlations with partial generalization of response; and generally high, reliable, and positive correlations with generalization of stimulus.[1] For both integral and partial values, generalization of response and generalization of stimulus yield negative correlations, the consistent but not always reliable occurrence of which seems to justify the differentiation of these types of generalization in this paper.

Variation with Learning.—It is of interest to note the progress of the various types of response with learning. As direct comparison is not possible, owing to the varying number of syllables and of learning trials in the four series, a modification of the Vincent-Kjerstad method substantially following that of Melton (4), was used to compare the results.

While it is recognized that some form of learning was evident in the partial and generalized responses, we shall follow the progress of learning as given by the integral discriminated responses at every trial. Accordingly, for every series, the integral discriminated responses were first smoothed by a moving average, and then averaged for number of syllables and number of subjects, to enable us to trace at every trial the learning in terms of a fraction of a syllable. Naturally, complete learning would be indicated by an integral discrimination of 1.00 syllable. Since our learning was not carried beyond a discrimination of 0.13 syllables, it is obvious that our findings are restricted to the initial stages of learning.

Due to limitations of space, the curves for the individual series are not presented. Although the learning curves of integral discrimination for all four series indicate a slight initial positive acceleration for about four or five trials, inspection, indicates that the most characteristic trend for all curves seems to be that of a straight line. It will be remembered that in this experiment learning was still in the initial stages and the characteristic deceleration of learning curves had not yet definitely appeared. Accordingly, to all four curves a straight line was fitted by the method of least squares. In general the fit is good, the standard error of estimate being, in terms of a fraction of a syllable, 0.0180, 0.0064, 0.0069, and 0.0054, for the 6-, 8-, 10- and 12-Series respectively. From the lines thus calculated it is seen that the learning rate (based upon the slope of

[1] As would be expected, these findings, based on the intercorrelations, confirm the conclusions reached from a study of the curves of the relation between degree of discrimination and generalization, in both this paper and the previous one.

the lines) varies inversely with the length of the series, for the equations of the lines, proceeding from the shortest to the longest series, are respectively; $Y = 0.0079 + 0.0152X$; $Y = 0.0155 + 0.0108X$; $Y = 0.0088 + 0.0084X$; and $Y = 0.0029 + 0.0073X$, where X indicates learning trial, Y indicates the fraction of syllable learned, and the coefficient of X indicates the slope of the line.[2]

Points of equal learning, (the same fraction of a syllable integrally discriminated) were then determined on these lines, at every fraction of a syllable from 0.03 to 0.13, to include nearly all the learning trials in each series. Thus an integral discriminated value of 0.13 syllable was found for the 6-Series at 8.1 trials; for the 8-Series, at 10.6 trials; for the 10-Series, at 14.6 trials, and for the 12-Series, at 17.5 trials. As each series was presented one and a half times for every syllable in the series, it is seen that all series were carried approximately to the same degree of learning, but that the learning was far from complete.

The equal learning points, based on the number of trials for each series required to learn successive fractions of a syllable, were used in the determination of average curves of discrimination and of generalization. As has been stated, this is substantially Melton's (4) procedure of using a time-per-unit-of-work method.

For each type of response (integral discrimination, integral generalization, partial discrimination, partial generalization, and inadequate response), scores made on successive trials for every series were smoothed by a moving average,[3] from which curves were plotted, but they are not presented in this paper. The curves for integral discrimination have already been described. In general the curves for the other types of response follow consistent trends from series to series, with the following exceptions. The curves of integral generalization of response are all negatively accelerated, with maximum values occurring at slightly different points some few trials before the end of the series. The partial generalization curves are of two types: for the shorter series they rise to a maximum about the middle of the learning, and then fall steadily; for the longer series they drop from the beginning and reach a minimum which they maintain for the last few trials.

On all of these curves of the different variables based on the smoothed averages the *points of equal learning* were determined, from which the corresponding values of the different variables were found

[2] This finding agrees with that of Melton (*op. cit.*) who found that the Kjerstad-Robinson law does not hold for nonsense syllables, when the time-per-unit-of-work method is used to determine progress of learning.

[3] The smoothing formula was $X_n = \frac{X_{n-1} + 2X_n + X_{n+1}}{4}$.

XUM

by linear interpolation. Thus, for instance, for each of the four series, the fraction of a syllable that was integrally generalized was determined at every point of equal learning. At each of these points the four values thus obtained were averaged to yield an average integral generalization curve dependent upon degree of learning. Curves were obtained in a similar manner for all the variables; they are presented in Fig. 3. From this figure we are able to see not only the average trend of the variables with learning, but also the proportionate amount at each stage of learning.

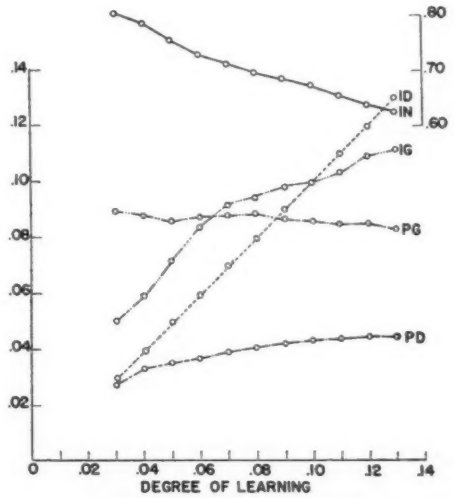


FIG. 3. Variation of types of response with degree of learning.

The ordinates are the response values in fractions of a syllable, averaged from all four series, as determined for points of equal learning. The abscissæ are the points of equal learning in terms of fractions of a syllable which is integrally discriminated. ID stands for Integral discrimination; IN, Inadequate values; IG, Integral generalization; PG, Partial generalization; PD, Partial discrimination. The ordinate scale on the right hand side is for inadequate values only.

In Fig. 3 the average learning curve for integral discriminated values appears as a straight line, which, of course, is to be expected, since it is derived from the four lines of best fit obtained from each series. Since the integral discrimination values were the basis from which learning was compared, it was felt that a mathematical smoothing was required in their case alone. Consequently the slight initial positive acceleration does not appear on this average curve. The other curves represent the values smoothed only by the moving average. But they too, of course, obscure some of the main trends. Thus, the point of inflexion at the maximum in all the integral gen-

eralization curves is not brought out because the maxima occur at slightly different points. The average curve of partial generalization falls continuously from the beginning of learning; and the points of inflexion on two of the actual curves do not appear. Generally speaking, however, the main tendencies of all variables with degree of learning are brought out by this figure.

It is apparent that the generalized responses, both integral and partial, are more frequent than discriminated responses during most of the learning to which this experiment was carried. However, the trend towards the end of the learning was for discrimination to increase and for generalization to diminish, for both partial and integral responses. Furthermore, for both discrimination and generalization, the integral responses are decidely more numerous than the partial responses at the end of the learning of this experiment. Inadequate responses are always more numerous than all the others, though they decrease steadily from the beginning of the learning.

Discussion

A tentative description of the progress of learning is here advanced, based on the results of the learning curves in this experiment. Owing to the low association values of the syllables and the rotation of the recitation order, the learning task was quite difficult. Learning was far from complete and was not spread uniformly over all the syllables. At the outset, those syllables which were easiest, due to their serial location, pronunciation, similarity, associative value, etc., were first selected by the subject for learning, probably without formulated intention or even definite awareness of such selection.

If we consider the various types of response to represent different stages of learning, of which the inadequate response is the initial stage and the integrally discriminated response is the terminal stage, we may state that, from almost the first trial, learning proceeds by stages, all of which are carried on simultaneously. At the outset of learning generalization tends to precede discrimination, that is, a syllable may be integrally reproduced before it is properly localized. Some syllables are only partially reproduced, and at all but the first few learning trials, they are less frequent than the integral syllables. All other syllables are in general disregarded; the subject finds that he cannot learn all at once and prefers to concentrate on the few that he can learn. Learning proceeds not so much by the selection of additional syllables for learning, as in the correct localization of the generalized responses, the integral reproduction of the partial responses, and the partial reproduction of the inadequate responses. With repeated trials, inadequate responses steadily diminish and integrally discriminate responses increase. Near the end of the

XUM

learning to which this experiment was carried the integral generalized responses reach their maximum and then diminish, because the integral responses are localized at a faster rate than the partial responses are integrally reproduced; for, according as the material selected for study passes through the various stages of learning, the partial responses become fewer in number.

Naturally this progress from stage to stage of learning is not uninterrupted and many integral responses may become partial or even inadequate for some few learning trials; thus, there is oscillation at the threshold of recall, as was pointed out by Hull (3). Moreover, recency is a decided factor in determining the type of response reproduced (7). As the recitation order is constantly changing, a syllable disregarded in all previous trials, in the sense that it does not show even partial learning, may abruptly appear as an integral response for a few trials and then become partially or inadequately reproduced for some trials thereafter. This fact was frequently observed in tabulating the results, especially when a difficult syllable occurring late in the learning order happened to appear at the beginning of one of the recitation trials. For that trial and for a few subsequent ones its reproduction was considerably improved.

So far we have been describing the reproduction of the response in an attempt to show its development with learning. The other variable, localization, also undergoes progressive changes with learning, although they are not so clearly revealed in this study. When the total responses from all the learning trials are taken into consideration, it is found that localization depends in part on its serial distance from the basic syllable—its gradient of generalization. The gradient seems to have a slight but reliable effect for integral values only; the correctly reproduced syllables are not randomly attached to any stimulus, for their localization tends to be most precise in the neighborhood of the basic syllable. The gradient for partial response is of questionable reliability and occurs at positions towards the end of the series. It may be that partial responses, not being so adequately reproduced as integral responses, are susceptible of localization only to the degree that they are less frequently placed in positions remote from the basic syllable.

Possibly of more importance in localization is the fact that generalization of stimulus varies directly with discrimination, while generalization of response tends, if at all, to vary inversely with discrimination. These relationships may be seen from the table of intercorrelations, Table III, and the figures showing the relationship between discrimination and generalization, Fig. 2 of this paper and Fig. 2 of the previous paper (7). Only a tentative explanation of these facts can be here advanced, and it is partly in terms of the

selective attitude of the subject and his tendency to concentrate upon certain syllables to the exclusion of others that are harder for him. It must be remembered that the stimuli need only to be recognized, but that the responses must be both reproduced and localized. When the subject cannot accurately localize a response, he is prone to assign it to those stimulus syllables which he most readily recognizes and discriminates; it is for this reason that the generalization value of a stimulus is prone to vary directly with its discrimination value, when the individual syllable is concerned. It seems also that the explanation suggested in the previous paper is of pertinence here. The response members of the pair, being less subject to inhibitory influences than the stimulus members, do not exhibit as much variability in generalization or discrimination as do the stimulus members, and hence the relationship between generalization and discrimination is much less clear-cut for response members.

Localization, therefore, depends upon at least two factors, the positional gradient and a factor intrinsic to the syllable itself. There is some indication that there is a progressive change in the relative influence of these factors with progress in learning, which may be obscured when only the total responses are considered in localization.

The results of this paper, both for reproduction and for localization, have been explained on the basis of a tendency by the subject, after the first trial or two, to select certain syllables for learning. This tendency implies a quantitative and selective set, not necessarily explicitly formulated, to concentrate on those syllables which can be brought within his span. As learning proceeds the selection is maintained on the same syllables to which others are gradually added. Sometimes fluctuations occur in the selective set, occasioned by changes in the recitation order which brings syllables, hitherto disregarded, into favored positions. The evidence for this explanation is largely indirect, and is gleaned chiefly from a study of the consecutive recitals of individual subjects, and their comments on their mode of learning.

Throughout learning, as was pointed out in a previous paper, intraserial effects are important and are most readily explained by some inhibitory hypothesis, such as those of Hull (**2**) or Foucault (**1**). It is felt, however, that intraserial inhibition does not account for all the phenomena in the progress of learning, and should be supplemented by some such explanation as that here advanced, based upon selective set.

The description of the progress of initial learning of nonsense syllables given in this paper is based upon group results. No attempt has been made to follow for individual learning the various stages here described. It is felt, however, on the basis of an ac-

quaintanceship gained from tabulation of the material, that such an analysis would not reveal facts much different from those here presented dealing with the gradual improvement in reproduction and localization which comes early in learning.

Summary

This paper deals with the initial stages of the learning of nonsense syllable lists by the paired associate method. Learning was not carried to completion in this experiment; however each series was carried approximately to the same degree of learning. In the beginning of this study certain questions were proposed; the following is a brief summary of the answers which have been found:

1. The relative proportions of discriminated and generalized responses decrease with length of list.

2. Generalized responses are more frequent than discriminated responses during most of the learning. Towards the end of the learning, discriminated responses increase while generalized responses decrease, and integral responses become more numerous than partial responses.

3. There is a slight gradient for serial position which is more pronounced for the shorter than for the longer series, and is more apparent for integral responses than for partial responses.

4. Discrimination varies directly with both integral and partial generalization for the stimulus; the variation is negative or zero for the response.

5. A tentative explanation of the results is advanced in terms of inhibitory effects within the series, and an unformulated set, on the part of the subjects, to select certain syllables, upon which they concentrate.

(Manuscript received April 21, 1943)

Bibliography

1. Foucault, M. Les inhibitions internes de fixation. *Année psychol.*, 1928, **29**, 92–112.
2. Hull, C. L., *et al.* *Mathematico-deductive theory of rote learning; a study in scientific methodology.* New Haven: Yale Univ. Press, 1940.
3. Hull, C. L. The influence of caffeine and other factors on certain phenomena of rote learning. *J. gen. Psychol.*, 1935, **13**, 249–273.
4. Melton, A. W. The end spurt in memorization curves as an artifact of the averaging of individual curves. *Psychol. Monogr.*, 1936, **47**, no. 212, 119–133.
5. Melton, A. W., & Irwin, J. M. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **33**, 170–204.
6. Peixotto, H. E. Intraserial inhibition as measured by reproduction. *J. exp. Psychol.*, 1942, **31**, 17–35.
7. Philip, B. R., & Peixotto, H. E. Generalization in the initial stages of learning nonsense syllables. I. Integral responses. *J. exp. Psychol.*, 1943, **33**, 50–63.

THE INFLUENCE OF SIZE OF TEST STIMULI, INTERPUPILLARY DISTANCE, AND AGE ON STEREOSCOPIC DEPTH PERCEPTION

BY LEONARD C. MEAD

Tufts College

Introduction

The present paper presents an analysis of the relation between the size of test stimuli and their 'decentration' (*i.e.*, disparity produced by lateral shift of one of the paired images) in a standard stereographic test of depth perception. This analysis was suggested by the casual observation, during a test of this type, that many subjects showed an inability to disregard the size cue in making depth judgments supposedly based on binocular retinal disparity. Although this confusion was to be expected from common knowledge of the cues utilized in making judgments of distance, the writer has been unable to find a specific experiment describing the relationship. Other data, gathered at the time of testing, will permit the presentation of intercorrelations between stereopsis score, interpupillary distance, and age.

There are numerous experiments in the literature in which the various cues to depth perception have been combined so as to augment or counteract each other, thereby leading to conclusions as to the relative strength or weakness of certain single or grouped cues. Much of the published work has been an attempt to rate the relative contribution of tactual and kinesthetic cues (from accommodation and convergence) as compared to the visual cues of size, perspective, binocular disparity, double images, etc. This topic has been reviewed by Carr (**3**), Woodworth (**19**) and Boring (**2**); and more recently it has been treated by Grant (**7**), who concluded that both accommodation and convergence play an equal and important role in determining the distance of near meaningful objects.

The present writer is concerned specifically with the interaction between retinal size and disparity cues when presented stereoscopically. Most of the previous experiments on this topic have varied line perspective, rather than overall size, in determining the relation between this factor and binocular disparity. Schriever (**14**) found that disparity always overcame perspective, but his paired photographs were made by moving the camera a distance of 12 cm. between pictures, thus approximately doubling the usual interpupillary

XUM

distance disparity. When shading, perspective, and superposition acted together, they afforded depth judgments appropriate to them rather than to the reversed disparity. Contrary binocular parallax was unable to overcome superposition alone, the impression sometimes being that of transparency.

Washburn and Wright (**17**) used pairs of lines on both sides of stereograms. Separation between pairs of lines was 1 cm. when no depth was present, and 5 mm. on one side when disparity was introduced so as to produce apparent depth. Size changes (perspective) were effected by varying the length of the lines which were either 4 or 5 cm. long. They also varied 'intensity' by controlling the darkness of the lines. Their general finding concerning the efficiency of these cues considered singly was that disparity was stronger than perspective which in turn was stronger than intensity. Data were also presented showing the effect of the various combinations of one or more of these factors, either acting alone or against other combinations.

Although Weber and Bicknell (**18**) used stereoscopic slides made by photographing objects of different sizes at different distances with unequal degrees of disparity, the judgments required by the instructions to their subjects were not directed toward the effect of size on distance perception. They were concerned with 'phenomenal regression to real size' which they found to occur in photographs, just as Thouless (**15**) found phenomenal regression toward the real size of objects.

Vernon (**16**) noted some effects of size and binocular disparity cues in her experiments in 'real' rather than 'stereoscopic' space. With the surrounding environment carefully controlled, she found during observations on 'binocular fusion sensations' that several observers reported using the size rather than the disparity cue. Other experiments involved the situation where the variable and standard stimuli were equally distant but of different sizes. Here she found no tendency for the subjects to judge the smaller stimuli farther and the larger stimuli nearer than the standard stimulus. Vernon concluded that other sensory and secondary data are operative.

In the experiment to be described below, preliminary data of which have been presented in abstract form by Mead and Kennedy (**11**), stimuli of four different sizes occur in a series of stereograms in which there is a decreasing amount of binocular disparity.

Procedure

The test used was the Stereometric Unit, DC 1–23, manufactured by the Keystone View Company. It consists of a series of 23 stereograms representing decreasing differences in binocular parallax. Each stereogram has three rows of ten characters (letters and numbers). One

member of each row is made to stand out stereoscopically in front by "decentration" with respect to the rest of the row. For the three rows on any card, decentration parallax (stereopsis angle) is defined by the amount of decentration and the position of the card on the bar of the Telebinocular. Fry (5) has calculated these angles for the various amounts of decentration of the 23 cards. A score of 'percent stereopsis' is obtainable from each card, '100 percent stereopsis' being defined as the angle of binocular parallax equal to a difference of 25 millimeters on the Howard-Dolman (Howard, 9) apparatus.

The ten characters of each row occur in one of four sizes. Size is designated in terms of the visual acuity required to recognize the number or letter. These four acuities are, respectively, from largest to smallest, 20/120, 20/90, 20/50, and 20/25. Instructions for administering the test recommend stating that "size of the letter is no indication of its 'distance.'"[2]

Sixty-seven male subjects, most of them undergraduates at Tufts College, took the test.[1] The only stipulation regarding their vision was that they did not wear glasses. A small percentage of test applicants was rejected because of inability to fuse the stimuli. The observers were allowed to move the card along the Telebinocular bar to the point of easiest fusion.[2] After obtaining comfortable fusion with Card 1, a subject was required to call off the letter or number in the top, middle, and bottom rows which seemed to stand out nearer to him than the rest of the row. The admonition that the size of a figure bore no relation to its apparent distance was emphasized at this point, and was repeatedly given throughout the remainder of the series. If a subject did not seem to 'get the point,' the experimenter coached him on Card 1. After study of the first card, the second was presented. The subject was eliminated if he was unable to discriminate correctly on the latter. All 67 subjects were then required to go through the remaining 23 slides. Testing time varied between 5 and 18 minutes. Responses were recorded on prepared answer sheets by the examiner, and a 'percent stereopsis' score was given the subject after he finished the test.

Other data gathered for each subject included his age and interpupillary distance, the latter being measured by the Shuron Pupillometer.

Results

The responses given by the subjects for each of the three rows of the 23 stereograms were tallied under two main categories, namely, 'correct' and 'incorrect.' Within these two categories, the responses were designated according to the size (visual acuity) of the character selected by the subject as appearing nearer to him than the rest of the row. The frequencies of selection of stimuli of various sizes are indicated in Table I. The symbols *T*, *M*, and *B* stand for Top, Middle, and Bottom rows. It will be noted that one column of this table, under the division 'Incorrect Responses,' is headed by a question mark. This means that some subjects, besides being unable to select the decentered character, gave as a response a letter or number which was actually not on the card.

In Table II are presented the percentages of correct and incorrect responses also designated according to size. No account is taken here, however, of the rows on each card, although the writer has separately made such an analysis. Just as in the preceding table,

[1] The assistance of Miss Ruth E. Clark, Dr. Arthur C. Hoffman, and Dr. John L. Kennedy in administering the tests is gratefully acknowledged.

[2] The Keystone View Company instructions say this is allowable, although they recommend, in two different instructions, setting the slide at 1.25 D. and 0.0 D. Such changes naturally change the calibration of the slides (Fry, 5).

there is no entry under at least one of the four sizes of 'Correct Responses' due, obviously, to the fact that there are only three rows on each stereogram.

The data of Table II are shown graphically in Figs. 1 and 2. Fig. 1 shows the percentage of correct 'nearer' judgments made by

TABLE I

FREQUENCY OF CORRECT AND INCORRECT RESPONSES MADE BY 67 SUBJECTS ANALYZED WITH RESPECT TO SIZE OF TEST STIMULI AND VERTICAL POSITION OF TEST STIMULI

Card Number		Correct Responses				Incorrect Responses				
		20/120	20/90	20/50	20/25	20/120	20/90	20/50	20/25	?
1	*T*	67				0	0	0	0	
	M		65			2	0	0	0	
	B				61	5	0	0	0	1
2	*T*		52			9	4	2	0	
	M				45	16	1	2	0	
	B	63				1	2	1	0	
3	*T*	58				1	1	5	1	1
	M			48		15	2	2	0	
	B	59				3	1	3	0	
4	*T*				29	30	0	6	0	2
	M			43		0	22	2	0	
	B	59				2	2	3	0	1
5	*T*			30		23	2	9	1	2
	M	61				1	1	2	1	1
	B	59				1	3	4	0	
6	*T*		39			22	0	6	0	
	M	57				5	5	0	0	
	B				25	33	3	6	0	
7	*T*		30			28	1	4	1	3
	M				10	43	7	5	2	
	B	59				5	1	2	0	
8	*T*	57				8	2	0	0	
	M				15	36	9	7	0	
	B		29			32	4	2	0	
9	*T*				14	35	11	6	0	1
	M		26			29	0	10	2	
	B			23		35	9	0	0	
10	*T*				16	37	6	7	0	1
	M	27				31	2	6	1	
	B			14		41	8	2	0	2
11	*T*				16	35	10	5	0	1
	M	51				9	5	1	1	
	B			13		39	7	7	1	
12	*T*				14	22	22	8	1	
	M			14		44	3	3	2	1
	B	52				3	6	5	0	1

TABLE I—*Continued*

Card Number		Correct Responses				Incorrect Responses				
		20/120	20/90	20/50	20/25	20/120	20/90	20/50	20/25	?
13	*T*		31			31	1	3	1	
	M				14	40	4	7	2	
	B			17		42	5	0	3	
14	*T*				15	38	2	11	1	
	M	47				6	6	6	1	1
	B			11		43	3	6	4	
15	*T*		22			38	0	5	2	
	M			20		39	4	1	3	
	B				8	31	3	13	0	2
16	*T*	41				11	3	10	0	2
	M				11	46	6	4	0	
	B			11		40	5	8	3	
17	*T*		21			37	3	5	0	1
	M			13		42	7	3	1	1
	B				5	47	10	0	1	4
18	*T*				6	45	2	7	7	
	M	30				17	12	6	2	
	B		15			41	4	7	0	
19	*T*				5	36	13	11	1	1
	M		6			51	1	7	1	1
	B	31				19	0	14	3	
20	*T*			8		40	11	6	2	
	M				3	42	11	11	0	
	B	34				17	3	10	3	
21	*T*	37				10	8	8	4	
	M			2		48	5	5	5	2
	B				4	42	11	8	2	
22	*T*				1	43	9	4	10	
	M	16				37	7	4	3	
	B			4		47	9	2	5	
23	*T*				0	41	5	14	5	2
	M		3			43	8	7	5	1
	B			2		41	14	5	0	5

the 67 subjects for the 23 slides as a function of size of the correct test stimuli. In cases where one of the four acuity sizes does not appear as the correct stimulus on a card, the curve representing this size is continued from the card of its last appearance to the card of its next appearance. It should be remembered that as the card number increases, the binocular disparity of the correct stimulus decreases. Despite certain irregularities in the curves of Fig. 1, the trend seems clear. With equivalent amounts of binocular parallax, the percentage of correct responses is directly related to the size of

TABLE II

PERCENTAGE OF CORRECT AND INCORRECT RESPONSES MADE BY 67 SUBJECTS ANALYZED WITH RESPECT TO SIZE OF THE TEST STIMULI

Card Number	Correct Responses				Incorrect Responses				
	20/120	20/90	20/50	20/25	20/120	20/90	20/50	20/25	?
1	100	97.0		91.0	3.5				.5
2	94.0	77.6		67.2	12.9	3.5	2.5		
3	87.3		71.6		9.5	2.0	5.0	.5	.5
4	88.1		64.2	43.3	15.9	11.9	5.5		1.5
5	89.6		44.8		12.4	3.0	7.5	1.0	1.5
6	85.1	58.2		37.3	29.9	4.0	6.0		
7	88.1	44.8		14.9	37.8	4.5	5.5	1.5	1.5
8	85.1	43.3		22.4	37.8	7.5	4.5		
9		38.8	34.3	20.9	49.3	10.0	8.0	1.0	.5
10	40.3		20.9	23.9	54.2	8.0	7.5	.5	1.5
11	76.1		19.4	23.9	41.3	10.9	6.5	1.0	.5
12	77.6		20.9	20.9	34.3	15.4	8.0	1.5	1.0
13		46.3	25.4	20.9	56.2	5.0	5.0	3.0	
14	70.1		16.4	22.4	43.3	5.5	11.4	3.0	.5
15		32.8	29.9	11.9	58.7	3.5	9.5	2.5	1.0
16	61.2		16.4	16.4	48.3	7.0	10.9	1.5	1.0
17		31.3	19.4	7.5	62.7	10.0	4.0	1.0	3.0
18	44.8	22.4		9.0	51.2	9.0	10.0	4.5	
19	46.3	9.0		7.5	52.7	7.0	15.9	2.5	1.0
20	50.7		11.9	4.5	49.3	12.4	13.4	2.5	
21	55.2		3.0	6.0	49.8	11.9	10.4	5.5	1.0
22	23.9		6.0	1.5	63.2	12.4	5.0	9.0	
23		4.5	3.0	0.0	62.2	13.4	12.9	5.0	4.0

the test stimuli. Thus the stimuli requiring the least visual acuity (20/120) are more often correctly perceived, for a given amount of decentration, than stimuli requiring any other acuity. The 20/90 size stimuli are the next most readily perceived, and the 20/50 sizes are for the most part intermediate in discriminability between the 20/90 sizes and the 20/25 sizes.

Fig. 2 presents the relation between the percentage of incorrect responses for the 23 slides as a function of size of the incorrect test stimuli. Unlike the curves of Fig. 1, where each curve is displaced from the others by an amount approximately proportional to the stimulus size, the curves of Fig. 2 show that the 20/120 stimuli receive a disproportionate emphasis as compared to the other sizes. Although the smallest stimuli (20/25) are least often incorrectly perceived as nearer, the two intermediate sizes (20/90 and 20/50) are equally often incorrectly selected as being nearer than the remainder of the row. It thus appears that if a subject is unable to discriminate a test stimulus on the basis of binocular parallax, he will choose the largest of the stimuli available for selection and will be relatively uninfluenced by smaller stimuli. In other words, if 'depth' is incorrectly perceived, test items smaller than the largest are not chosen in direct proportion to their size.

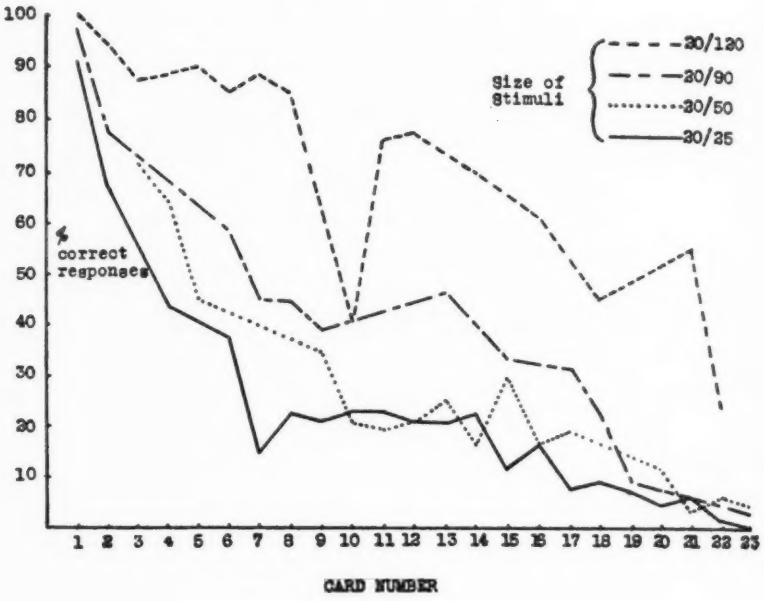

100
90
80
70
60
50
40
30
20
10
% correct responses
Size of Stimuli
20/120
20/90
20/50
20/25
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
CARD NUMBER

FIG. 1

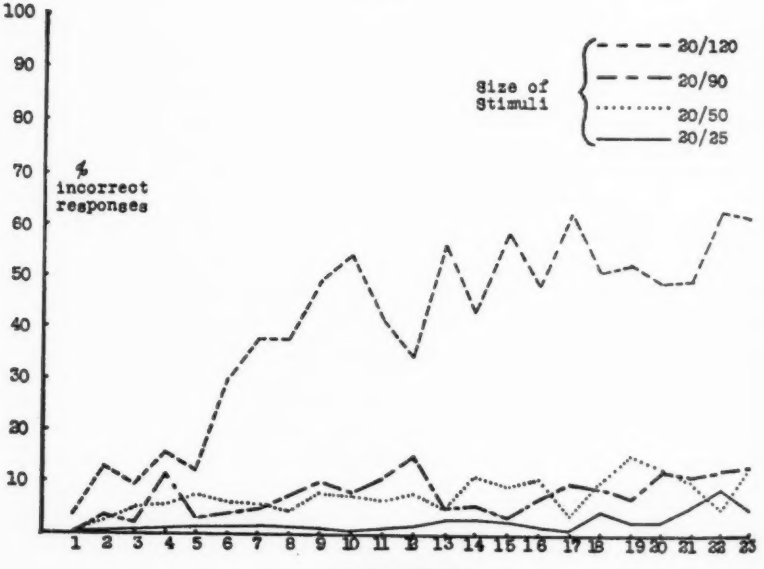

100
90
80
70
60
50
40
30
20
10
% incorrect responses
Size of Stimuli
20/120
20/90
20/50
20/25
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
CARD NUMBER

FIG. 2

Although it is not presented in this paper, the writer has made an analysis similar to those shown graphically in Figs. 1 and 2 in which the different rows on the slides received separate treatment. The trends shown in the above figures were maintained regardless of whether the stimuli occurred in the top, middle, or bottom row of the stereograms. Thus the relationship pointed out above between sizes of test stimuli for equivalent amounts of stereopsis angle holds regardless of vertical position of the stimuli in the visual field.

The average age of the subjects was found to be 19.9 years (S.D. = 2.6), and average interpupillary distance was 63.2 millimeters (S.D. = 2.9). A stereopsis score on the Keystone slides was obtained by counting the total number of responses which were incorrect on the basis of binocular parallax; thus, a perfect score would be 0 and the poorest score would be 69. The average score was found to be 41.1 (S.D. = 11.9). Although this procedure for scoring is not the one recommended by the manufacturers of the test, it utilizes the data from the whole series of cards. The standard instructions permit stopping of the test when failures begin.

Pearson product-moment correlation coefficients were computed for the above three sets of data. The correlation between age and stereopsis score was −.382, between age and interpupillary distance −.148, and between stereopsis score and interpupillary distance −.182. In order to be significantly different from chance (Lindquist, **10**) at the 1 percent level of confidence, with this number of cases, r must exceed .317. Thus, the only reliable correlation coefficient obtained is that of −.382 between age and number of incorrect responses. In other words, there was a slight but reliable tendency for the older subjects to get a better score on the Keystone test.

Discussion

The results presented above are not considered by the present writer to settle the question of the relative influence or importance of the size cue as compared to the disparity cue in depth perception. As a matter of fact, it seems highly probable that 'depth' seen stereoscopically on a stationary stereogram is a unique type of perception, differing markedly from the usual perceptual reaction to natural tridimensional visual space (Higley, **8**). It is well recognized that mere size offers an ambiguous cue to apparent distance (Carr, **3**), since a retinal image of constant size may be produced by larger and larger objects at progressively increasing distances. Even binocular disparity, undoubtedly one of the most important cues to depth, does not 'force' the depth effect (Woodworth, **19**). In almost any group of observers, there will generally be a small percentage who, after

fusion of the images, do not see the apparent depth when objects are presented stereoscopically. Woodworth (**19**) points out that "binocular disparity is a stimulus, but the depth effect is a response." A phenomenological response could be flatness and disparity rather than plastic (relief) and single vision.

Numerous recent investigators and writers (Boring, **2**) have emphasized the importance of dyanmic principles in visual depth perception. Thus Vernon (**16**) points out that mere disparity does not operate to produce depth according to geometrical rules, but that the 'set' of the observer and the formal structural characteristics of the visual field play important determining roles. Minami (**12**) also emphasizes the importance of 'attitude.' Schlosberg (**13**), although writing on the illusion of depth from single pictures, appropriately points out that "depth perception cannot be considered as a mere addition and subtraction of factors of differing importance." He suggests that the quantitative observation of the influence of single factors can sometimes be made better by eliminating other factors, rather than by employing the common method of holding all other factors constant. Not only may certain cues have different mathematical functions when considered singly, but there may be an interaction between cues when presented in combination. Thus, the apparent size of different objects at different distances in a photographic view may change in a stereoscopic view involving disparity and other depth cues (Carr, **3**; Gehrke, **6**; Weber and Bicknell, **18**).

The considerations presented above, suggesting that binocular depth perception depends on a highly complex and dynamic integration of many cues working together, plus the 'set' and 'attitude' of the observer, and the 'surround' and knowledge of the matrix in which the observations occur, show that conclusions regarding the absolute effects of size and disparity cues in stereoscopy are unwarranted. Rather, the contribution of the present experiment is to show the importance of these two factors in a common, although specific, test situation. Although 'stereopsis' tests of the type used above are useful, caution should be exercised in the interpretation of a subject's ability to discriminate in terms of binocular parallax when the test objects are of different sizes. The curves shown in Fig. 1 indicate that the 'stereopsis threshold' is also a function of the size of the test stimuli. Percentage of correct judgments of 'nearer' for a given stereopsis angle varies as the size of the objects. If one is interested in measuring the sensitivity of a subject to binocular parallax, it would appear more appropriate to keep all the test objects at a constant size.

Another feature of the present experiment is the lack of significant correlation between interpupillary distance and stereopsis test score.

XUM

In view of the complicated nature of depth perception, it is not surprising to find that a single aspect of the depth response, namely, the amount of disparity produced by the separation between the eyes, does not correlate highly with the total response. Although Banister and Blackburn (1) concluded that wider average interpupillary distance was found among a group of 258 Cambridge University undergraduate students who were rated 'good' in athletics, they point out that the rating of athletic proficiency was affected by halo, bias, and lack of information. Clark and Warren (4) found no correlation between interpupillary distance and depth perception as measured by the Howard-Dolman test. Neither was the average interpupillary distance significantly different between a group of athletes and an unselected group.

Although the correlation between age and stereopsis test score was reliable in this experiment, it is too low to be of general importance.

Summary and Conclusions

A standard test of stereoscopic depth perception was administered to 67 male subjects whose average age was 19.9 years and whose average interpupillary distance was 63.2 mm. The test consisted of 23 stereograms with three rows of 10 characters on each card. Decentration of one character in each row made this member appear nearer than the remainder of the row. The degrees of binocular retinal disparity on any card were the same, but the stereopsis angle decreased with increasing card number. Since the various numbers and letters on the 23 cards were of four different sizes, it was possible to determine the effect on the judgment of apparent depth of size of the test stimuli at different amounts of binocular parallax. This analysis was performed for both correct and incorrect judgments.

The results permit the following conclusions:

1. For a given degree of binocular retinal disparity, the percentage of correct 'nearer' judgments is directly related to the size of the test objects.

2. For a given degree of binocular retinal disparity, the percentage of incorrect 'nearer' judgments is not directly proportional to the size of the test objects. Instead, the largest stimuli elicit a disproportionately greater number of responses than the smaller sizes of stimuli.

3. The conclusions cited under (1) and (2) above hold regardless of the vertical position of the rows of test stimuli in the visual field.

4. No significant intercorrelations were found between number of incorrect responses, age, and interpupillary distance. The correlation between age and stereopsis score was reliable but not large in magnitude.

5. Since the discriminability of different amounts of binocular parallax is related to test stimulus size, a test which is designed to measure this function should preferably utilize objects of constant size throughout the test.

6. The perception of apparent depth in a test of this type is considered to be a highly specific and unique type of response, and only partially related to the natural type of depth and distance perception in free visual space.

(Manuscript received May 21, 1943)

References

1. Banister, H., & Blackburn, J. M. An eye factor affecting efficiency at ball games. *Brit. J. Psychol.*, 1931, **21**, 382–384.
2. Boring, E. G. *Sensation and perception in the history of experimental psychology.* New York: Appleton-Century Co., 1942. Pp. 263–311.
3. Carr, H. A. *An introduction to space perception.* New York: Longmans, Green & Co., 1935. Pp. 195–288.
4. Clark, B., & Warren, N. Depth perception and interpupillary distance as factors in proficiency in ball games. *Amer. J. Psychol.*, 1935, **47**, 485–487.
5. Fry, G. A. The measurement of the threshold of stereopsis. Mendenhall Laboratory of Physics, Ohio State Univ., Columbus, Ohio.
6. Gehrke, E. Neue Versuche zur physiologischen Optik. *Z. tech. Physik.*, 1940, **21**, 292–295.
7. Grant, V. W. Accommodation and convergence in visual space perception. *J. exp. Psychol.*, 1942, **31**, 89–104.
8. Higley, H. R. Stereopsis and depth perception. *Columbia Optomet.*, 1941, **15**, No. 57, 2–4.
9. Howard, A. J. A test for the judgment of distance. *Amer. J. Ophthalm.*, 1919, **2**, 656–675.
10. Lindquist, E. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin Co., 1940. Pp. 210–214.
11. Mead, L. C., & Kennedy, J. L. The influence of size of test stimuli in the measurement of stereopsis. *Psychol. Bull.*, 1942, **39**, 502. (Abstract.)
12. Minami, H. (An experimental study on the dominant factors of visual space perception.) *Jap. J. Psychol.*, 1940, **15**, 153–180. (Seen only in *Psychol. Abstr.*)
13. Schlosberg, H. Stereoscopic depth from single pictures. *Amer. J. Psychol.*, 1941, **54**, 601–605.
14. Schriever, W. Experimentelle Studien über stereoskopisches Sehen. *Z. Psychol.*, 1925, **96**, 113–170.
15. Thouless, R. H. Phenomenal regression to the real object. *Brit. J. Psychol.*, 1931, **21**, 339–359.
16. Vernon, M. D. The perception of distance. *Brit. J. Psychol.*, 1937, **28**, 1–11; 115–149.
17. Washburn, M. F., & Wright, C. Studies from the Psychological Laboratory of Vassar College. LXVIII. The comparative efficiency of intensity, perspective, and the stereoscopic factor in producing the perception of depth. *Amer. J. Psychol.*, 1938, **51**, 151–155.
18. Weber, C. O., & Bicknell, N. The size constancy phenomenon in stereoscopic space. *Amer. J. Psychol.*, 1935, **47**, 436–448.
19. Woodworth, R. S. *Experimental psychology.* New York: Henry Holt & Co., 1938. Pp. 651–683.

XUM

A CUMULATIVE GRAPHIC WORK-RECORDER

BY O. H. MOWRER

Harvard University

Graphic recorders of conventional design, whether of the kymographic or continuous tape varieties, have a common limitation when used for cumulative work recording: they permit only one excursion of the recording element across the writing surface. The recording element can, of course, be re-set manually so as to allow for the beginning of a new excursion; but this procedure takes time and may introduce artifacts into the record. Or, the re-setting may be made to occur automatically, but this requires so complicated a mechanism as to be impractical for most purposes. The usual expedient is simply to reduce the magnification factor to such an extent that no amount of work that is likely to occur within the observation period can produce more than a single excursion. This means that slight responses are likely to be quite illegible and that the whole record lacks detail. Through the use of an unusual relationship between the writing element and the record proper—which, so far as the author is aware, has not previously been described—, all these difficulties can be circumvented.

The usual arrangement employed in graphic recorders provides that the recording surface shall be moved past the writing element at a constant speed and that the variable work shall be performed upon the writing element. In the apparatus shown in Fig. 1, this relationship is reversed: the writing element is driven at a constant speed and the writing surface is moved by the variable work.[1] Fig. 2 shows a sample record obtained by connecting the pulley located at the right end of the apparatus (see Fig. 1) by means of a light belt to another, smaller pulley which was mounted on a shaft supporting a treadmill arrangement which has been described elsewhere (1). This record shows the work performed by a rat which was required to revolve the treadmill at least 18° every 60 sec.; if the rat failed to make such a response, it received an electric shock which lasted until the task was performed. The record is read from left to right and shows the distribution of treadmill activity during a 20-min. test period. The points at which the shock was applied (note the learning) are checked (∧). The horizontal components of the record

[1] The author is indebted to Mr. Ralph Gerbrands and to Dr. S. S. Stevens for a number of useful suggestions concerning the construction of this apparatus.

indicate no movement of the treadmill; the ascending components indicate speed and amount of treadmill activity. Each complete vertical excursion of the record represents eight revolutions of the treadmill. If made on coordinate paper, such a record can be directly quantified. Otherwise the use of scales of the type shown in Fig. 2 is indicated. However, the main function of records of this kind is not so much that of quantification (counters of various kinds will serve this purpose) as it is that of providing a detailed, compact picture of the *total pattern* of work performed during a given period.

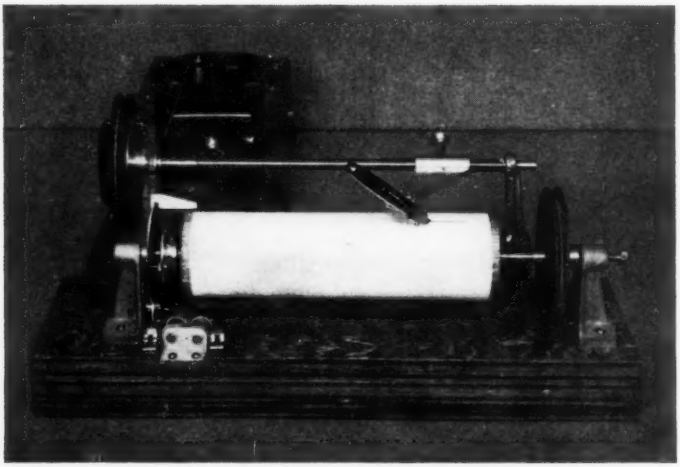


FIG. 1. A cumulative graphic work-recorder especially designed for situations in which considerable detail (relatively high magnification) and unlimited recording capacity are both desirable.

The amount of work which can be thus recorded is obviously very great. If desired, the various parts of such a record can, of course, be cut apart and reassembled to form a single, continuous graph.

In making records of the kind just described, the clutch at the left end of the cylinder carrying the record (see Fig. 1) is disengaged from the ratchet wheel so that the cylinder turns freely. In the particular apparatus here described, this cylinder is made of solid wood and mounted on a steel shaft which has frictionless point bearings. If great lightness were essential, the cylinder could, of course, be made of cardboard or thin sheet-metal. The inertia and momentum of the wooden cylinder have not, however, seemed excessive for any purpose to which the recorder has thus far been put. The paper on which the record is made can be fastened to the cylinder either by sealing the overlapping edges with Grippit (which can be

ruptured without tearing the paper when the record is removed) or by means of a rubber band looped over each end of the cylinder after the paper has been wrapped around it. The paper (ordinary $8\frac{1}{2}$ by 11 in. coördinate sheets) can be changed with the cylinder in

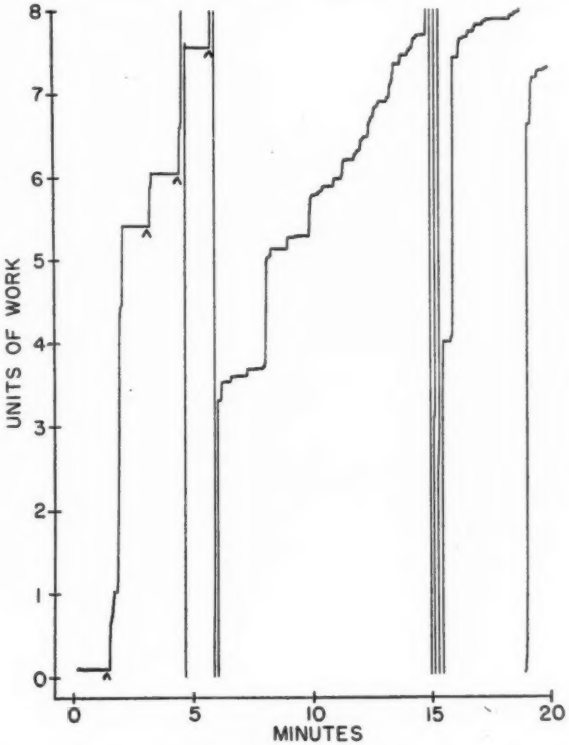


FIG. 2. Illustrative record made with apparatus shown in Fig. 1, under conditions described in the text. Note that this is a single record, to be read continuously from left to right and from bottom to top. Regard the top of each vertical component of the record as continuous with the bottom of the next vertical component to the right. By reference to the scale of Units of Work shown at the left, it is possible to determine the total amount of work performed (in this case, slightly over 71 units) within the observation period (20 minutes).

position; or, by the loosening of the set-screw at the right end of the cylinder shaft, the whole cylinder can be removed for this purpose.

During the making of a record, the writing element (a common fountain pen) moves at a constant speed from the right to the left side of the apparatus. As it nears the end of its excursion, the writer automatically lifts off the paper as it rides up on the rest-arm, which can be seen in Fig. 1. Shortly thereafter the writer-

carriage stops moving, due to the fact that there are no threads on the drive-rod at the extreme left end. This arrangement means that the constant-speed motor which revolves the carriage drive-rod can be allowed to continue running indefinitely after a record is completed, without danger of jamming the writer-carriage at the end of an ex-

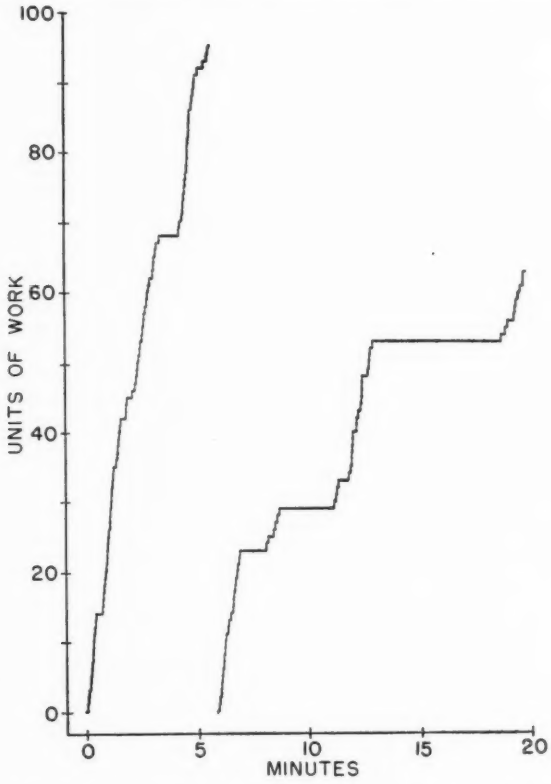


FIG. 3. Record showing accumulation of discrete events. If a single curve is desired, the latter part of the record can be moved upward so as to be continuous with the first part.

cursion. To return the carriage to the starting position, all that is necessary is to retract the small plunger whose threaded end normally presses against the drive-rod and slide the carriage to the right. Variation in the speed at which the carriage moves can be obtained by varying the size of the motor pulley and the drive-rod pulley.

In addition to its use in recording activity of the kind represented in Fig. 2, this apparatus can also be employed to register a succession

of work events, or responses, of a discrete, unitary nature. If the clutch, mentioned above, is engaged with the ratchet-wheel which is to be seen at the extreme left in Fig. 1, and if a constant torque is put on the pulley at the right by means of a string and a light weight, each operation of the escapement mechanism permits the cylinder to revolve 6° (*i.e.*, one tooth of the ratchet-wheel). This escapement can be operated either mechanically, by means of the small lever that can be seen projecting into the foreground in Fig. 1, or electrically, by means of the small electro-magnet located nearby.

Fig. 3 shows a record of the extinction of a Skinner bar-pressing habit (**2**), obtained with the apparatus functioning in the manner just described. In the original record it is easy to count the 160 separate responses which occurred during the 20-min. test period.

Although simple to construct and adaptable to a number of different uses, the apparatus here described is limited in that it cannot be used as a polygraph. Since the work is performed on the whole writing surface, multiple writing units would simply produce parallel records. Even a 'time line' is not possible, but since the rate at which the writer moves across the paper is constant, a time line can be later added to a record, as in Fig. 2 and Fig. 3. The distinctive feature of the device is that it permits the recording of an almost unlimited amount of work within a fixed period of time in such a form as to provide a compact, easily interpreted picture of the pattern and distribution of the activity.

(Manuscript received May 31, 1943)

References

1. Mowrer, O. H., & Miller, N. E. A multi-purpose learning-demonstration apparatus. *J. exp. Psychol.*, 1942, 31, 163–170.
2. Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century Co., 1938.